

**New catalytic strategies for DNA and RNA alkylation using  
rhodium(II) and copper(I) carbenes – a versatile tool for applications  
in chemical biology**

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

**Kiril Tishinov**

aus Bulgarien

Basel, 2015

**Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf  
Antrag von:**

Prof. Dr. Dennis G. Gillingham

Prof. Dr. Florian P. Seebeck

Basel, 9. Dezember 2014

**Prof. Dr. Jörg Schibler**

Dekan der Philosophisch-  
Naturwissenschaftlichen  
Fakultät

*I would like to thank everyone who helped me throughout my doctoral studies in the past four years – my supervisor, Prof. Dennis Gillingham, the whole Gillingham research group, all fellow PhD students, PostDocs, and colleagues at the Department of Chemistry, the University of Basel, all my friends and family.*



## Table of contents

List of abbreviations	vii
<b>Chapter 1. Introduction</b>	
<b>1.1. Strategies for nucleic acid modification</b>	<b>1</b>
<b>1.1.1. Pre-synthetic approaches for nucleic acid modification</b>	<b>1</b>
<b>1.1.1.1. Chemical approaches for pre-synthetic nucleic acid modification</b>	<b>4</b>
<b>1.1.1.2. Enzymatic and hemienzymatic approaches for pre-synthetic nucleic acid modification</b>	<b>7</b>
<b>1.1.2. Post-synthetic approaches for nucleic acid modification</b>	<b>10</b>
<b>1.2. Rhodium(II)-catalyzed X-H insertion reactions</b>	<b>18</b>
<b>1.3. Copper(I)-catalyzed X-H insertion reactions</b>	<b>24</b>
<b>1.4. Copper(I)-catalyzed azide-alkyne cycloaddition</b>	<b>26</b>
<b>Results and discussion</b>	
<b>Chapter 2. Nucleic acid alkylation with Rh(II) carbenoids</b>	<b>29</b>
<b>2.1. Alkylation of single-stranded DNAs and RNAs</b>	<b>29</b>
<b>2.2. Regioselectivity of the alkylation. Determination of the alkylation sites: tandem MS and NMR experiments</b>	<b>38</b>
<b>2.3. Scope of <math>\alpha</math>-diazocarbonyl substrates and conditions for nucleic acid alkylation</b>	<b>43</b>
<b>2.4. Mechanistic rationale for the Rh(II)-carbenoid nucleic acid modification</b>	<b>44</b>
<b>2.5. Structure selectivity of the Rh(II)-catalyzed alkylation: experiments with double-stranded DNAs and RNAs</b>	<b>48</b>
<b>2.6. Nucleic acid labeling applications</b>	<b>51</b>
<b>2.7. Conclusions and future directions</b>	<b>56</b>
<b>Chapter 3. Copper(I)-catalyzed insertion reactions in aqueous media</b>	<b>58</b>
<b>3.1. Alkylation of nucleic acids. Optimization of the reaction conditions and scope with single- and double-stranded DNAs and RNAs</b>	<b>58</b>
<b>3.2. Alkylation of simple arylamines in aqueous media</b>	<b>64</b>
<b>3.3. Tandem auto-catalytic Cu(I)-catalyzed azide-alkyne cycloaddition/N-H insertion. Biomolecular labeling applications</b>	<b>67</b>
<b>3.4. Conclusions and future directions</b>	<b>73</b>

<b>Chapter 4. Direct reversal of DNA alkylation with AlkB dioxygenase from <i>E.coli</i></b>	<b>74</b>
<b>4.1. Direct alkylation of mononucleoside-5'-monophosphates</b>	<b>77</b>
<b>4.2. Expanding the substrate scope of AlkB: direct enzymatic removal of bulky N-alkyl groups from nucleoside-5'-monophosphates</b>	<b>77</b>
<b>4.3. Conclusions and future directions</b>	<b>80</b>
<b>Chapter 5. Experimental</b>	<b>81</b>
<b>5.1. General</b>	<b>81</b>
<b>5.2. HPLC</b>	<b>82</b>
<b>5.3. Chemical synthesis</b>	<b>83</b>
<b>5.3.1. Monoalkylation of small anilines with copper(I) carbenoids derived from Dz1 in aqueous media</b>	<b>90</b>
<b>5.3.2. Auto-catalytic tandem CuAAC/N-H insertion with small anilines</b>	<b>95</b>
<b>5.3.3. Direct alkylation of 2'-deoxynucleoside-5'-monophosphates</b>	<b>97</b>
<b>5.4. Oligonucleotide synthesis and purification</b>	<b>100</b>
<b>5.5. Nucleic acid alkylation with rhodium(II) carbenoids</b>	<b>100</b>
<b>5.5.1. NMR Characterization of d(TCT), d (TAT) and d(TGT) and the products of alkylation with Dz1</b>	<b>107</b>
<b>5.6. Nucleic acid alkylation with copper(I) carbenoids</b>	<b>107</b>
<b>5.6.1. Cu(I)-catalyzed oligonucleotide modification using <math>\alpha</math>-diazocarbonyl compounds in the presence of proteins</b>	<b>118</b>
<b>5.6.2. Auto-catalytic tandem CuAAC/N-H insertion with nucleic acids</b>	<b>118</b>
<b>5.7. AlkB expression and purification</b>	<b>121</b>
<b>5.7.1. General</b>	<b>121</b>
<b>5.7.2. Subcloning of the alkB gene from <i>E.coli</i> LB21</b>	<b>122</b>
<b>5.7.3. AlkB overexpression and purification</b>	<b>124</b>
<b>5.7.4. Determination of delakylation activity</b>	<b>125</b>
<b>References</b>	<b>126</b>

## Abbreviations

pABSA – p-Acetamidobenzenesulfonyl azide

Asp – L-Aspartate

ATP – Adenosine-5'-triphosphate

Boc – tert-Butoxycarbonyl

bp – Base pair

BSA – Bovine serum albumin

BTT – Benzylthio-1H-tetrazole

CuAAC – Copper-catalyzed azide-alkyne cycloaddition

DBU - 1,8-Diazabicyclo[5.4.0]undec-7-ene

DMAP – 4-(*N,N*-Dimethylamino)pyridine

DMSO – Dimethyl sulfoxide

DMT – 4,4'-Dimethoxytrityl

DTT – DL-Dithiothreitol

Et<sub>3</sub>N – Triethylamine

dNTP – 2'-Deoxynucleoside triphosphate

ddNTP – 2',3'-Dideoxynucleoside triphosphate

FISH – Fluorescent *in situ* hybridization

His – L-Histidine

HPLC – High-performance liquid chromatography

IPTG - Isopropyl β-D-thiogalactopyranoside

KPi – potassium phosphate buffer

MES – 2-(*N*-morpholino)ethanesulfonic acid

MMT – 4-Monomethoxytrityl

MS – Mass spectrometry

MS/MS or MS<sup>2</sup> – Tandem mass spectrometry

MTase – Methyltransferase

Ni-NTA – Ni-Nitrilotriacetic acid resin

NMR – Nuclear magnetic resonance

NTP – Nucleoside triphosphate

OSPS – Oligonucleotide solid-phase synthesis

PCR – Polymerase chain reaction

Rh<sub>2</sub>(OAc)<sub>4</sub> – Rh<sub>2</sub>(μ-OOCCH<sub>3</sub>)<sub>4</sub>

SDS PAGE – Sodium dodecylsulfate polyacrylamide gel electrophoresis

TCEP - tris(2-Carboxyethyl)phosphine

TdT – Terminal 2'-deoxynucleotidyl transferase

THPTA - tris(3-Hydroxypropyltriazolylmethyl)amine

TMSOTf – Trimethylsilyl triflate

Tris - 2-Amino-2-(hydroxymethyl)-1,3-propanediol

pTsCl – p-Toluenesulfonyl chloride

UTP – Uridine-5'-triphosphate



## Chapter 1. Introduction

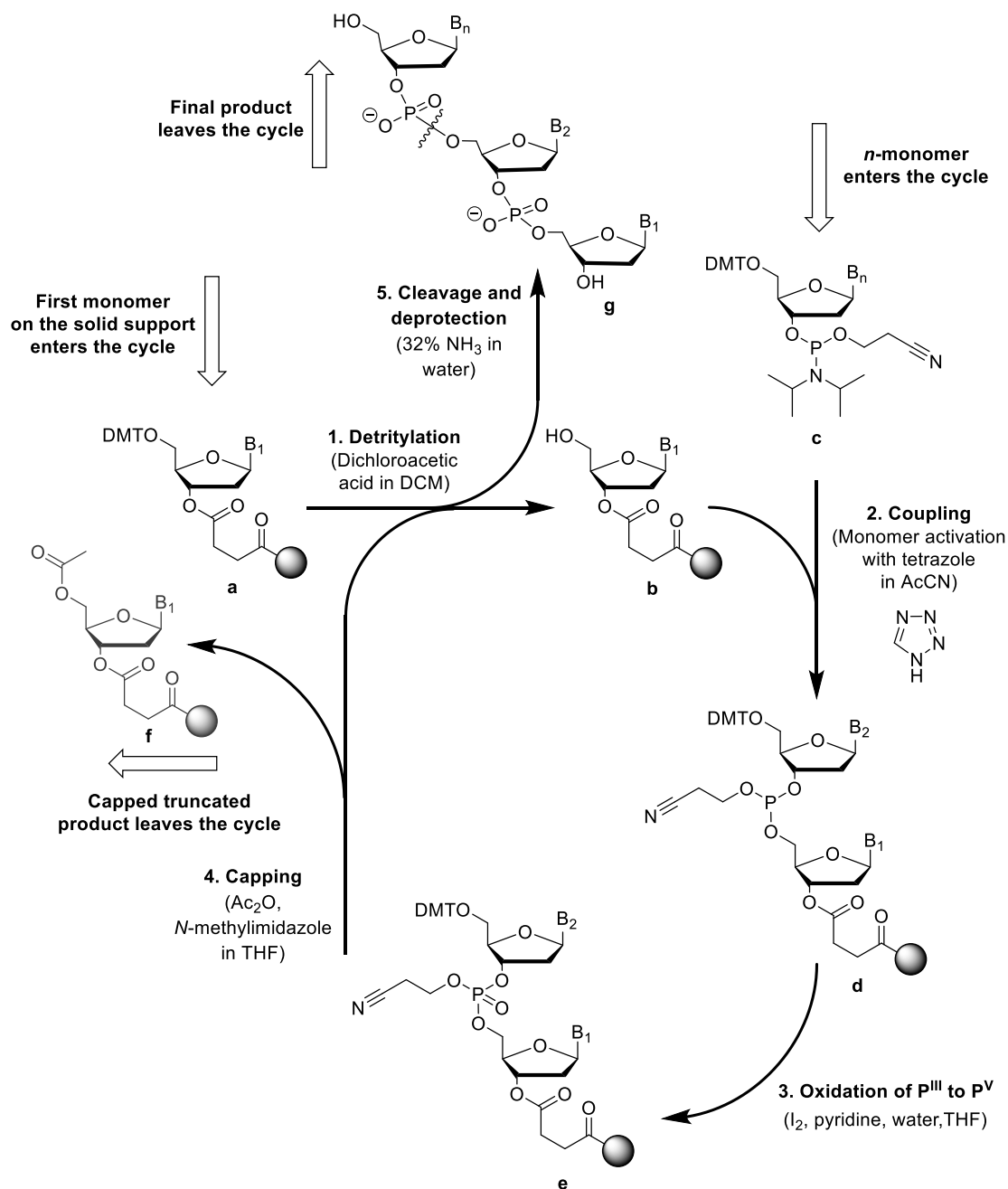
### 1.1. Strategies for nucleic acid modification

The field of biological and biomedical research relies heavily on strategies for biomolecular modification in order to label and expand the functional potential of the target structure. The development of methods for DNA and RNA tailoring is a key step not only to understand their biological role, but also to perturb their function for diagnostic and therapeutic use. The number of labels per oligonucleotide molecule and their exact location depends strongly on the respective application. In some examples for fluorescent *in situ* hybridization (FISH) oligonucleotides with only a single fluorescent label are used.<sup>1</sup> The construction of molecular beacons for real-time PCR<sup>2,3</sup> and cassette-labeled primers for DNA-sequencing<sup>4</sup> requires at least one donor/acceptor pair (two modifications) per DNA molecule, with a precisely defined distance in between for proper fluorescence quenching in the former case and efficient fluorescence enhancement in the latter. There are also cases where densely functionalized nucleic acids are needed, often with several contiguous modified bases, like some DNA architectures for biomedical and nanotechnology applications.<sup>5,6,7</sup> Such decorated DNAs and RNAs can be accessed by two general strategies. The modifications can be introduced pre-synthetically by using modified monomers for the nucleic acid assembly or post-synthetically by directly modifying the target DNA or RNA molecules.

#### 1.1.1. Pre-synthetic approaches for nucleic acid modification

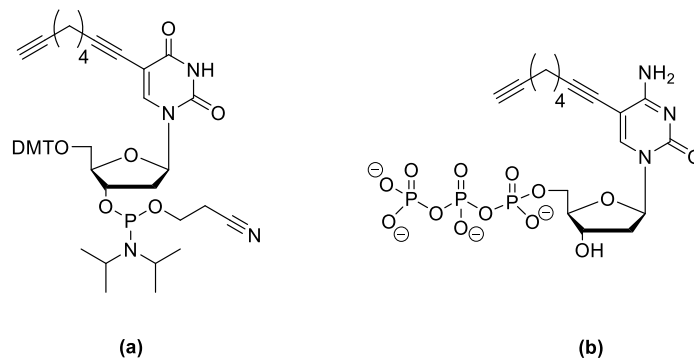
The nucleic acid assembly itself can be done in a number of ways. The most widely used chemical approach for the preparation of short and medium-sized oligonucleotides (up to 200 bases) is the automated oligonucleotide solid-phase synthesis (OSPS) using phosphoramidite chemistry.<sup>8</sup> An outline of the method is presented in Scheme 1. The first monomer enters the cycle already attached to the solid support (Scheme 1, **a**). It is then detritylated under acidic conditions to reveal the 5'-OH group of the growing chain (**b**). The second monomer unit is added by coupling the respective phosphoramidite **c** after activation with a weak acid (tetrazole). This is followed by oxidation of the phosphonium species **d** to a phosphonate **e**, and capping the unreacted 5'-OH groups in order to avoid elongation of failure chains. Any truncated products **f** will not react any further and are effectively removed from the synthetic cycle. Intermediate **e** can either be subjected to second coupling cycle or cleaved from the solid support and deprotected under basic conditions to afford the final product **g**. By this method the monomer units are joined in a step-wise fashion, elongating the newly-formed chain in a 3'-5' direction and a modification can be introduced at a specified position by the use of the respective tailored phosphoramidite precursor. PCR, chain extension and ligation are the main

enzymatic approaches for incorporation of pre-synthetically modified monomers into DNA, in this case NTPs.<sup>6,7,9,10,11,12</sup>



**Scheme 1** | Oligonucleotide solid-phase synthesis. (1) Detritylation with dichloroacetic acid in DCM reveals the 5'-hydroxyl group of the growing chain. (2) The 5'-O-protected-3'-phosphoramidite monomer is coupled to the 5'-hydroxyl group of the growing chain after protonation by tetrazole. (3) The phosphonium species **d** is oxidized to the phosphonate **e**. (4) Unreacted 5'-hydroxyl groups are 'capped' with an acylating reagent to prevent them from further reacting in subsequent coupling cycles. The growing chain can either be subjected to a second cycle of elongation (detritylation, coupling, etc.) or cleaved and deprotected under aqueous basic conditions (most often 32 % (v/v) ammonia in water for DNA or ammonia-ethanol for RNA) to afford the final product (5).

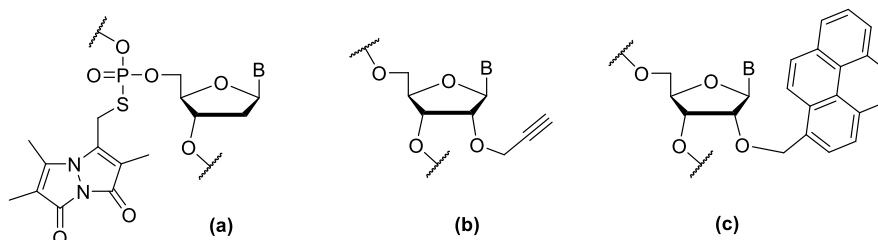
They are often unable to promote efficient coupling of some sterically demanding substrates, but are particularly valuable for preparation of longer DNAs. Two examples of modified monomers for introduction of an alkyne function in DNA - a phosphoramidite for OSPS (a) and a nucleoside triphosphate for PCR (b) - are shown below.



The preparation of modified RNAs is a far more complex task due to the greater susceptibility of the RNA molecules to chemical damage and enzymatic degradation, as well as the limited chain lengths (a little over a hundred bases) that can be accessed by OSPS.<sup>13</sup> Such cases often necessitate the use of more complex multistep synthetic strategies.

The label itself can be placed either terminally, at the 3'- or 5'-end hydroxyl groups, or inside the nucleic acid strand. In principle, an internal modification can be introduced at the phosphate backbone, on the ribose or on the nucleobase. One example of a phosphate modification is the introduction of multiple fluorophores on a phosphorothioate DNA as described by Conway and McLaughlin (Scheme 2a).<sup>14</sup> This approach, however, is not commonly utilized, since it perturbs the helix conformation and duplex stability, and interferes with most enzymatic reactions.<sup>15</sup>

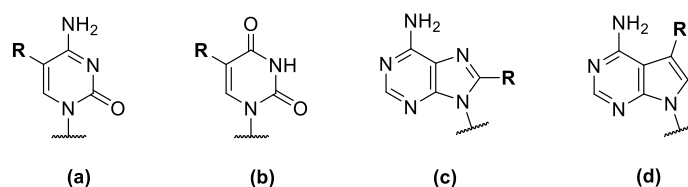
Functionalization of the ribose 2'-hydroxyl group has proven fruitful for label attachment in several cases, as substitution at that site does not greatly affect the duplex stability. This strategy was used by Wagenknecht et al. to introduce an alkyne function for 'click' chemistry attachment of azide dyes (Scheme 2b),<sup>16</sup> and by the group of Yamana for preparation of a 2'-intercalator-labeled detection system specific for RNA (Scheme 2c).<sup>17</sup>



**Scheme 2** | Modifications on the phosphate (a) and the 2'-hydroxyl group of the ribose ring (b, c).

When functionalizing the nucleobase in pyrimidine analogues C-5 is the preferred position for introduction of the modification, as it does not take part in Watson-Crick interactions and fits well in the major groove of the double helix with stabilities of the modified duplexes almost identical to those of the non-modified counterparts (Scheme 3a and b).<sup>6,10</sup>

Modifications on the purine bases, e.g. at C-8 (Scheme 3c), are generally avoided, as they disturb the overall stability of the double helix. When DNA polymerases are used for their incorporation such modification can interfere with their activity and significantly decrease the labeling efficiency.<sup>10,11</sup> The use of certain analogues, such as modified at C-7 7-deazaadenine (Scheme 3d), could alleviate this problem.<sup>18</sup>



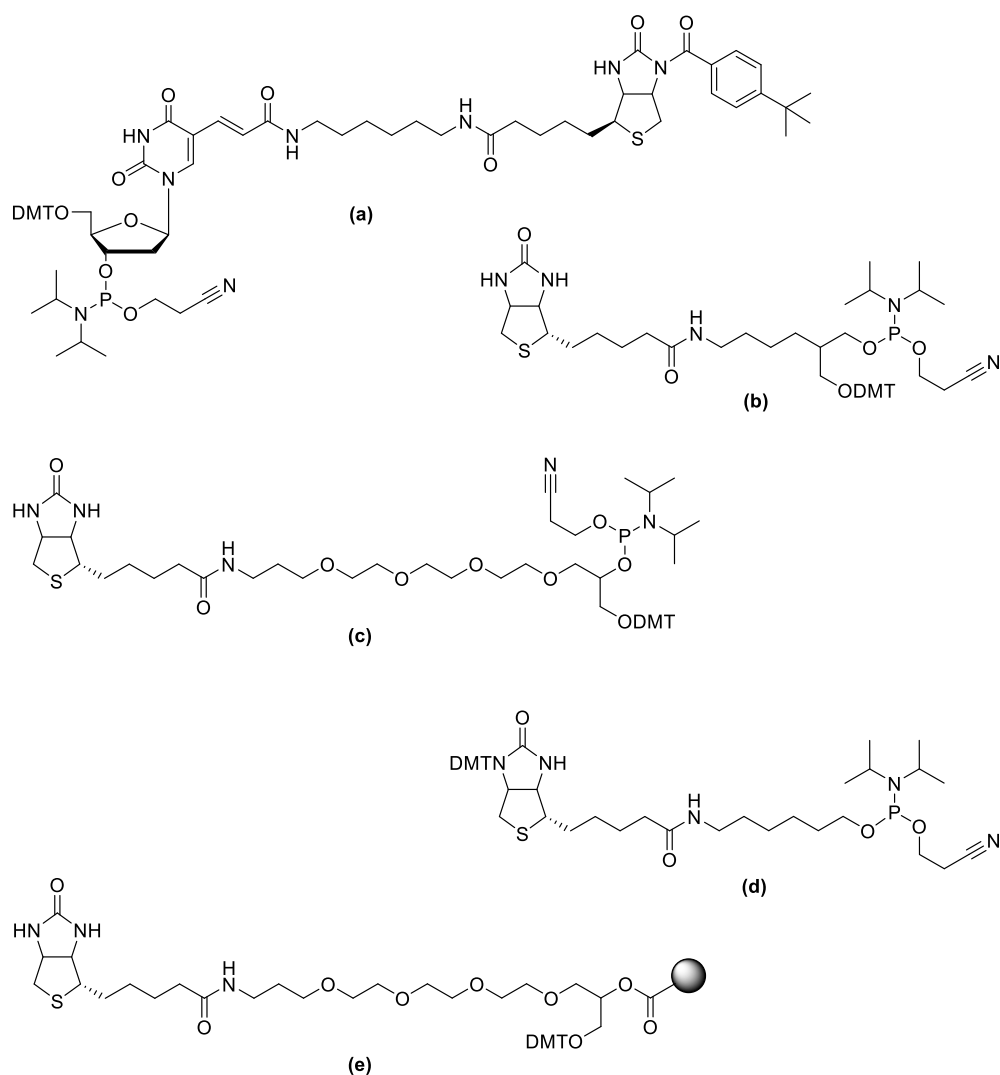
**Scheme 3** | Modification on the nucleobases: C-5 in pyrimidine analogues (a, b), C-8 in adenine (c), and C-7 in 7-deazaadenine (d)

### 1.1.1.1. Chemical approaches for pre-synthetic nucleic acid modification

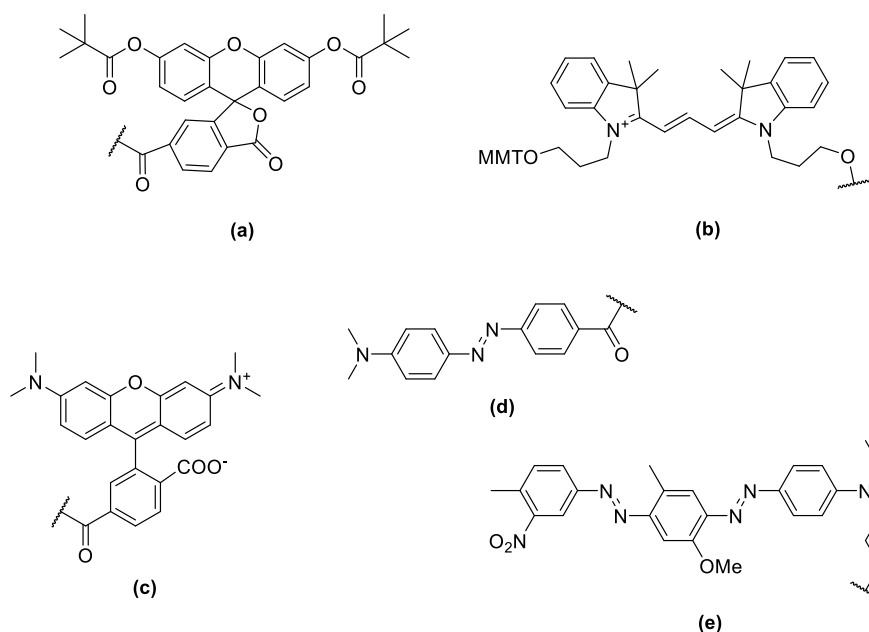
Oligonucleotide chemical synthesis is in principle quite flexible concerning the modifying group steric bulk and provides an access to a broader variety of attached tags than the enzymatic methods. It also allows the use of monomers that are considerably different than the natural nucleosides. Scheme 4 illustrates several examples for commercially available monomer units for the introduction of a biotin moiety by OSPS. In example **a** the tag is introduced at the C-5 of a thymine base through a hexyl diamine spacer arm. The *tert*-butylbenzoyl protecting group on the biotin is removed during the ammonium hydroxide deprotection step. Derivatives **b** and **c** are examples of non-nucleoside phosphoramidite monomers with different length and polarity of the biotin spacer, both suited for intrastrand incorporation. Example **d** is termed for 5'-termination, as it lacks a DMT-protected primary hydroxyl group for further chain elongation. The derivative in example **e** is used for 3'-labeling, as it is directly attached to the solid support. Removal of the DMT-group reveals a free primary hydroxyl group for attachment of the next unit.

A particularly important group of labels are different dye molecules – fluorophores and quenchers (Scheme 5). As discussed earlier they have a variety of practical applications and can be readily introduced by direct attachment to the respective monomers for oligonucleotide synthesis. Fluorescein-based dyes (a) are one of the most widely used fluorescent tags. The respective

phosphoramidites for both 3'- and 5'-end labeling, as well as for internal incorporation are commercially available for most of them. A second group of fluorescent probes, the cyanine dyes (b), are used mainly for real-time PCR and FISH analysis<sup>19</sup> and are introduced at one end of the chain, as internally they cause destabilization of the nucleic acid duplex. Rhodamine tags (c), used both as fluorophores and as quenchers, as well as dabcyI (d), and the Black Hole Quenchers<sup>TM</sup> (BHQ<sup>TM</sup>s, e) are three other examples of commonly applied dye molecules. The BHQ<sup>TM</sup>s have recently gain popularity, as they display high quenching efficiency and broad spectral overlap, which makes them suited for a number of advanced applications in modern genomics and diagnostics.

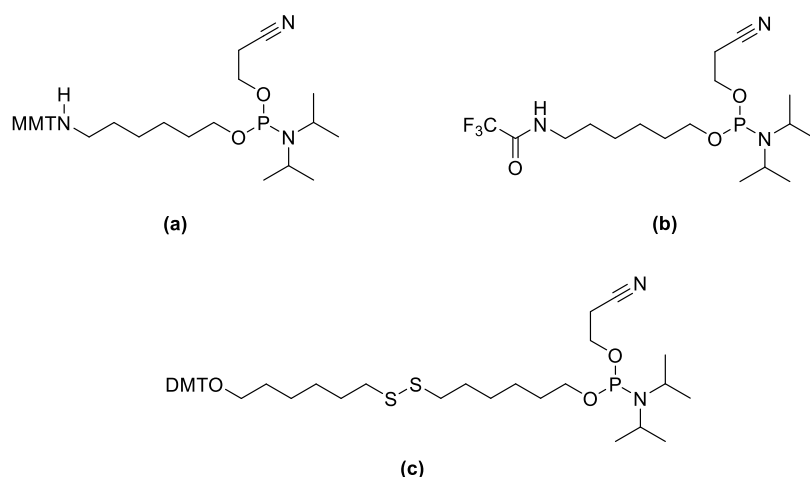


**Scheme 4** | Examples for commercially available phosphoramidites for introduction of biotin moieties.



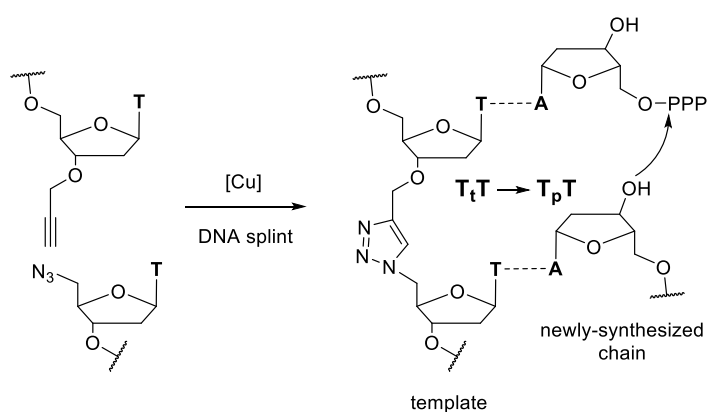
**Scheme 5** | Examples for commercially available dyes for direct introduction in OSPS. Fluorescein (a), Cyanine 540 (b), TAMRA (5-Carboxytetramethylrhodamine) (c), Dabcyl (d), and BHQ-1TM (e).

A practical strategy to chemically introduce a variety of new functional groups into DNAs and RNAs is through monomers bearing a specific tag, which after OSPS and deprotection is used to further decorate the oligonucleotide molecule. Often discussed as a post-synthetic modification strategy,<sup>15</sup> such approach should rather be classified as a pre-synthetic method, since the primary modification handle is installed already before the assembly of the nucleic acid. Masked amino or thiol groups (Scheme 6), that after deprotection can be modified with suitable electrophilic reagents (4-nitrophenyl and N-hydroxysuccinimide esters, maleimides), are employed quite commonly.



**Scheme 6** | Examples for commercially available phosphoramidites for introduction of nucleophilic groups for post-synthetic tagging. Masked 5'-terminal amino groups with acid-labile (a) and base-labile protection (b). Introduction of a thiol group protected as a disulfide (c).

The introduction of terminal alkynes for CuAAC is also a convenient approach for pre-synthetic chemical tagging. It has recently gained broad popularity due to the reliability and simplicity of the CuAAC and the compatibility of the alkyne and the azide groups used with a wide variety of biologically relevant functionalities. Besides the already discussed contributions from the Carrel<sup>5</sup> and Wagenknecht<sup>16</sup> groups, a particularly interesting example for pre-synthetic introduction of such tags comes from the group of Tom Brown, who have developed a ‘click’-able biocompatible linker for ligation of DNA strands.<sup>20</sup> 3'-Alkyne and 5'-azide terminated single-stranded DNAs, prepared by OSPS, have been efficiently ligated, affording products of up to 300 bases in length (Scheme 7). The products were read through with high fidelity during PCR providing the full-length amplicons. In addition, a gene containing the triazole linker was shown to be functional in *E. coli* cells.



**Scheme 7** | A biocompatible triazole DNA-linker developed by Brown et al.<sup>20</sup> Single-stranded DNA chains terminated by a 3'-alkyne and 5'-azide, are efficiently ligated by CuAAC using a bridging DNA template (DNA splint). The triazole linker in the new chain used as a template for a PCR-amplification is read through correctly.

### 1.1.1.2. Enzymatic and hemienzymatic approaches for pre-synthetic nucleic acid modification

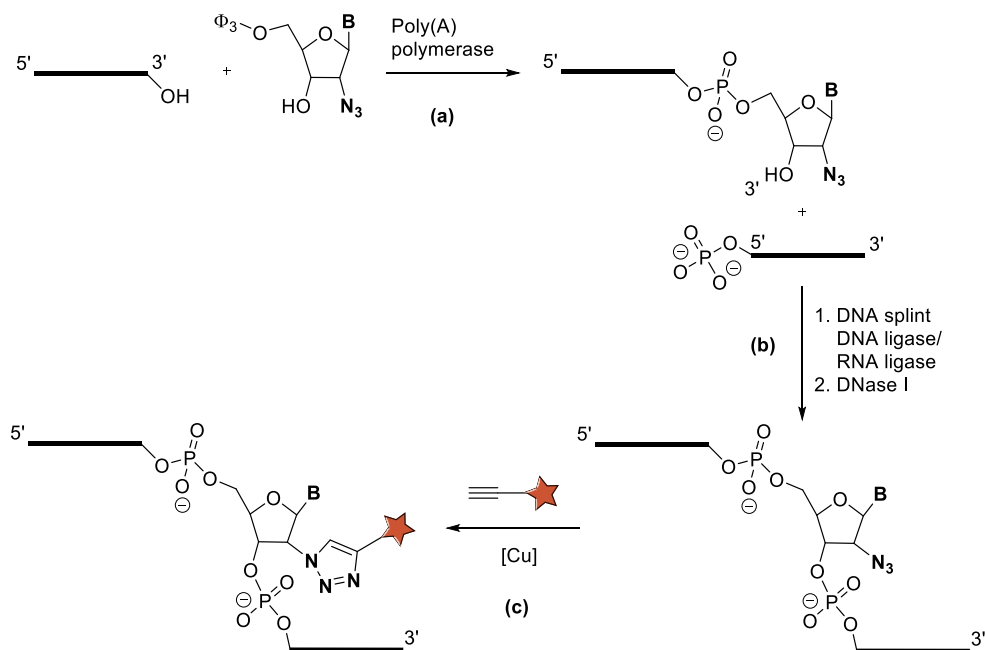
PCR and chain extension techniques with chemically tailored NTPs are the main enzymatic approaches for incorporation of modified nucleotides in DNA. Several examples from the Carrell group demonstrate the utility of the PCR method for the incorporation of alkyne-tagged monomers for subsequent tagging *via* CuAAC.<sup>6,7,12</sup> DNA 3'-end labeling with terminal deoxynucleotidyl transferase (TdT) and fluorescently-tagged NTPs is an example for a chain-extension technique.<sup>21</sup> TdT is a template-independent DNA polymerase that can extend a 3'-chain end with twenty to over a hundred bases. For some applications the attachment of just a single fluorophore per DNA molecule is required and in such cases a tagged ddNTP terminator is used.<sup>22</sup> The efficiency of incorporation, however, depends strongly on the structure of the fluorescent moiety.<sup>23</sup>

Direct enzymatic RNA 3'-end labeling can be done with poly(A) polymerase as described by Martin and Keller.<sup>24</sup> They have demonstrated the addition of dNTPs and ddNTPs to RNA for formation of



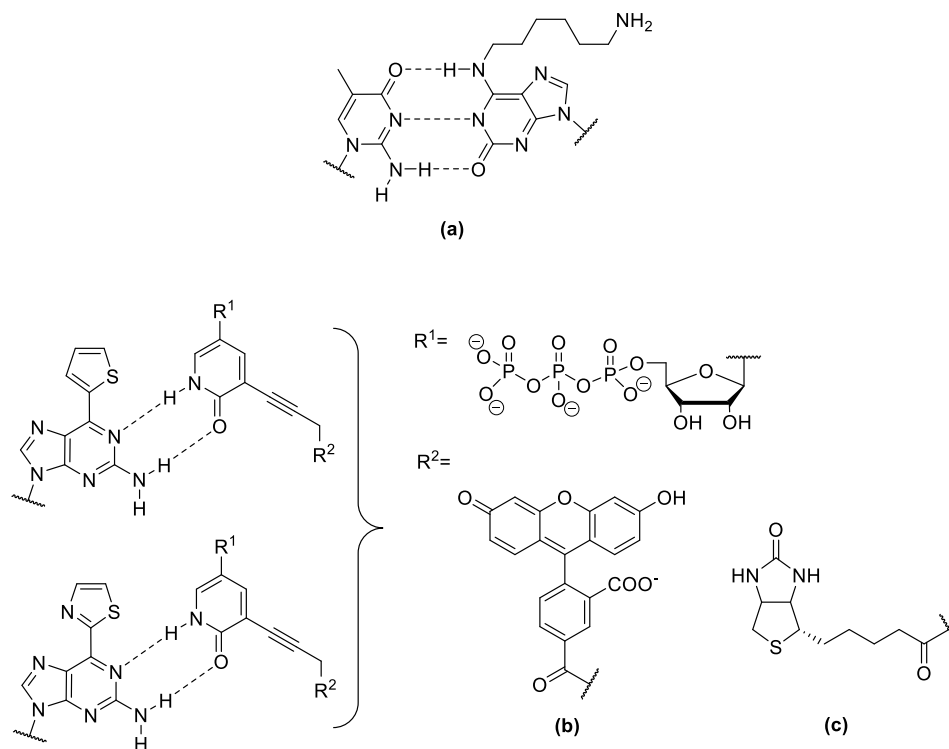


azidonucleoside-5'-triphosphate (a), followed by ligation of a second RNA to the new 3'-end (b) ultimately allowed the installation of an internal Alexa Fluor 647 tag by CuAAC (c).



**Scheme 10** | General scheme for introduction of an internal label in RNA using a poly(A) polymerase (a), ligation to a second RNA chain (b), and CuAAC with a fluorescently-tagged alkyne (c).

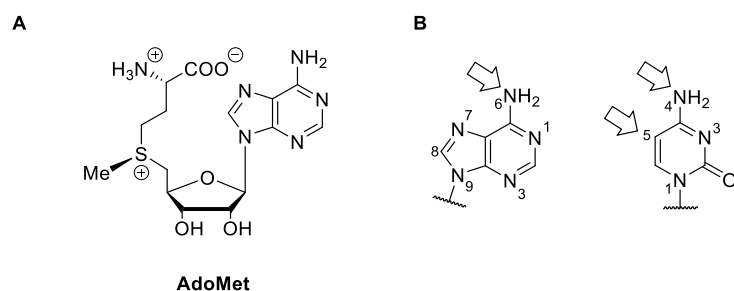
Modified RNAs can also be synthesized by using unnatural base pair systems. An unnatural base is installed on a DNA template and then used to direct the incorporation of a second unnatural base, uniquely complementary to it. An early example for this approach is the introduction of N<sup>6</sup>-(6-aminohexyl)isoguanosine in RNA by specifically pairing it to 5-methylisocytosine in the DNA template (Scheme 11a).<sup>27</sup> Hirao and co-workers have later expanded this methodology for the installation of fluorescent<sup>28</sup> and biotin tags<sup>29</sup> by pairing 2-amino-6-(2-thienyl)purine and 2-amino-6-(2-thiazolyl)purine to the labeled 2-oxo(1H)pyridine base (Scheme 11b and c).



**Scheme 11** | Modified RNA synthesis by using unnatural base pairs: 5-methylisocytosine paired with N<sup>6</sup>-(6-aminohexyl)isoguanosine (a), and 2-amino-6-(2-thienyl)purine or 2-amino-6-(2-thiazolyl)purine paired with 2-oxo(1H)pyridine for introduction of fluorescent (b) and biotin tags (c).

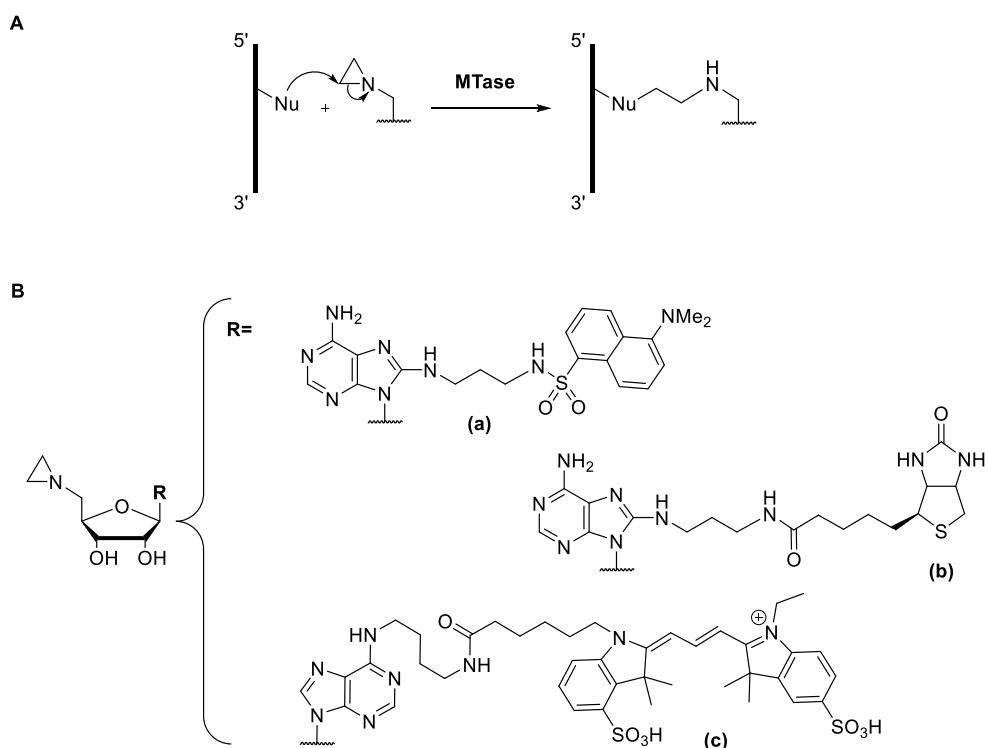
### 1.1.2. Post-synthetic approaches for nucleic acid modification

The modification of already synthesized DNA and RNA molecules is intuitively more direct and should provide access to structures of a greater complexity, especially in terms of chain length. One of the very few and most important **enzymatic techniques** for post-synthetic DNA and RNA modification involves the use of methyltransferases. Methyltransferases (MTases) are enzymes catalyzing the highly specific transfer of methyl groups from the ubiquitous cofactor *S*-adenosyl-L-methionine (AdoMet, Scheme 12A) to various positions in DNA or RNA (Scheme 12B). In the same manner MTase can operate with a variety of modified cofactors, allowing for sequence-specific DNA and RNA tagging.

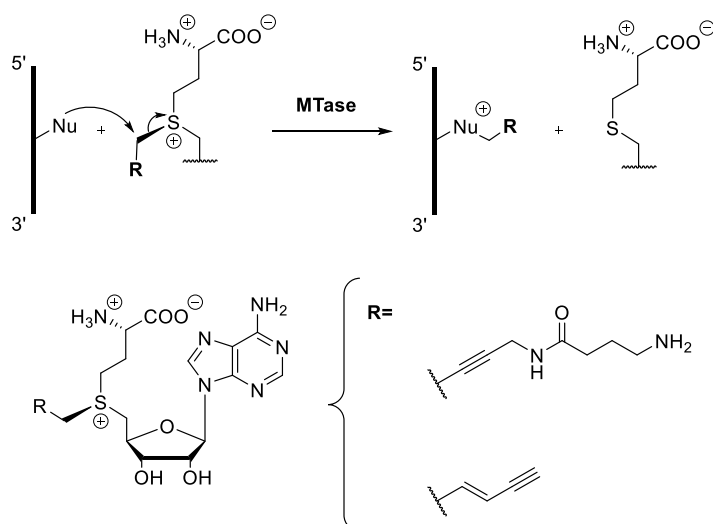


**Scheme 12** | Nucleic acid methylation, catalyzed by MTases. Structure of the AdoMet cofactor (A); possible alkylation sites depending on the MTase specificity: N-6 of adenine, N-4 and C-5 of cytosine (B).

The method was first established by the group of Weinhold for MTase from *Thermus aquaticus* (*M.TaqI*) and included substitution of the *S*-alkylmethionine moiety of the cofactor with an aziridine group for covalent attachment of the entire molecule to the target (Scheme 13A).<sup>30,31,32,33</sup> It was applied for plasmid DNA labeling with dansyl<sup>30,31</sup> and cyanine fluorophores,<sup>33</sup> and biotin,<sup>32</sup> installed on the cofactor's adenine base (Scheme 13B). The approach was further developed by modification of the *S*-alkyl group on the methionine sulfur alone for group transfer without attachment of the entire cofactor (Scheme 14).<sup>34</sup> Functionalization with terminal alkynes for example would allow further decoration of the target structure by using CuAAC, e.g. with fluorescent labels.<sup>35</sup>



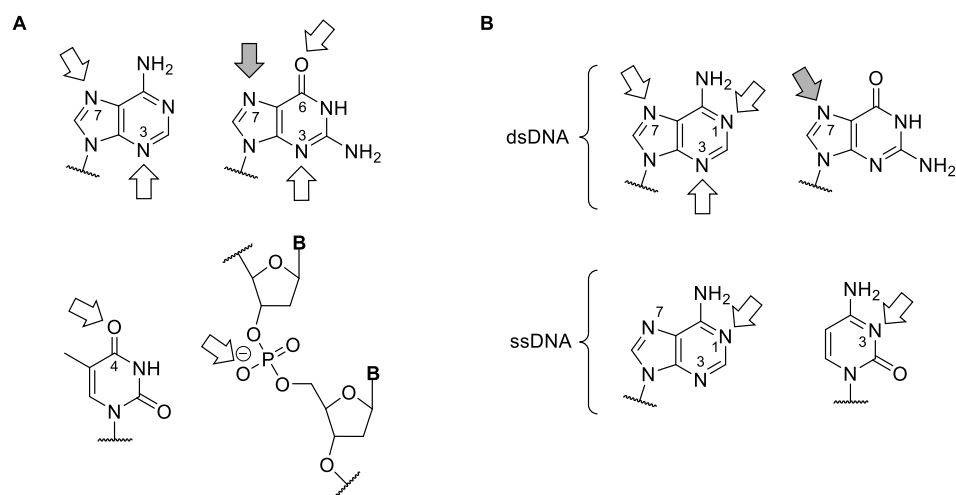
**Scheme 13** | Nucleic acid modification with an aziridine-modified cofactor. Mechanistic base: an attack by a nucleophile from the nucleic acid on one of the aziridine carbons leads to a ring opening and covalent attachment of the entire cofactor (A). Structure of modified cofactors for introduction of a dansyl (a), biotin (b), and a cyanine tag (c).



**Scheme 14** | Nucleic acid modification with a modified AdoMet cofactor, allowing sequence-specific group transfer.

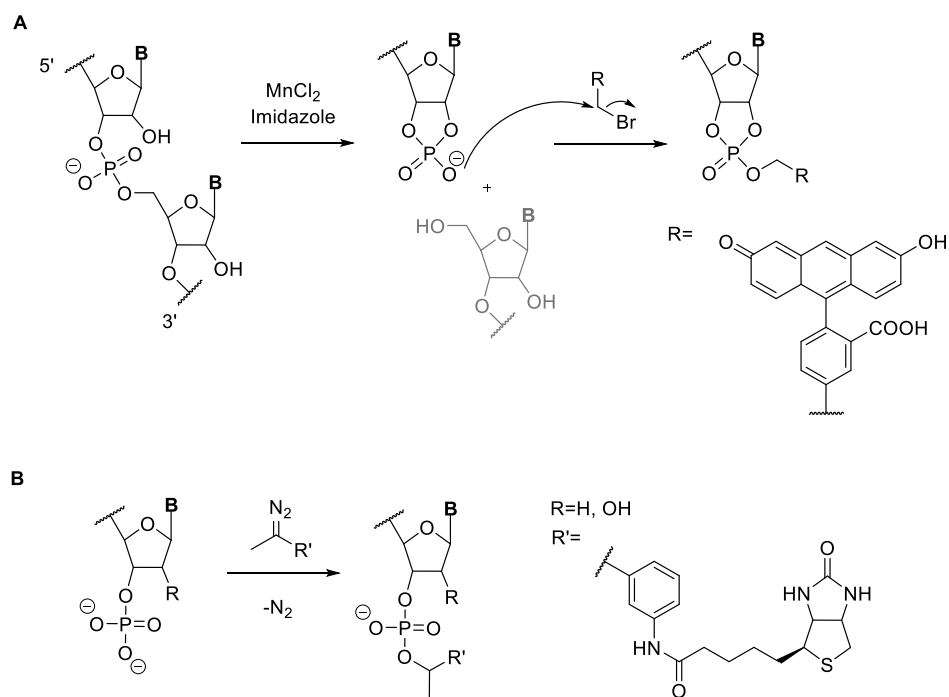
DNAs and RNAs can be post-synthetically modified by **chemical approaches**. In general, such approaches involve the use of activated electrophiles, which react with nucleophilic groups of the ribophosphate backbone or the nucleobases. Implementation of simple  $S_N1$  and  $S_N2$  alkylating reagents like alkyl halides and nitrosoureas is usually avoided, except for some specific applications. The methods utilizing these reagents suffer from low selectivity and require large excesses of the alkylating reagent. They, however, are quite valuable for establishing the relative reactivity of the different nucleic acid elements towards electrophiles and have helped correlating the action of a variety of carcinogens with the nucleic acid damage caused.<sup>36</sup> Mapping the potential alkylation sites can also assist the development of new DNA and RNA modification strategies by pointing out possible reaction targets.

$S_N1$  type alkylation reagents (*N*-methyl-*N'*-nitrosourea, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine) target O-6 in guanine, and to a lesser extent O-4 in thymine, as well as the backbone phosphates and N-3 of adenine and guanine (Scheme 15A).<sup>37</sup> In contrast,  $S_N2$  alkylating reagents (dimethyl sulfate, methylmethane sulfonate and methyl iodide) attack almost only nitrogen nucleophiles, especially N-1, N-3 and N-7 of adenine in single-stranded DNAs, and N-1 of adenine and N-3 of cytosine in double-stranded DNAs (Scheme 15B). It should be noted that N-7 of guanine is the primary target of both types of reagents, and is also the most frequently occurring nucleic acid alkylation damage found in double-stranded DNAs.



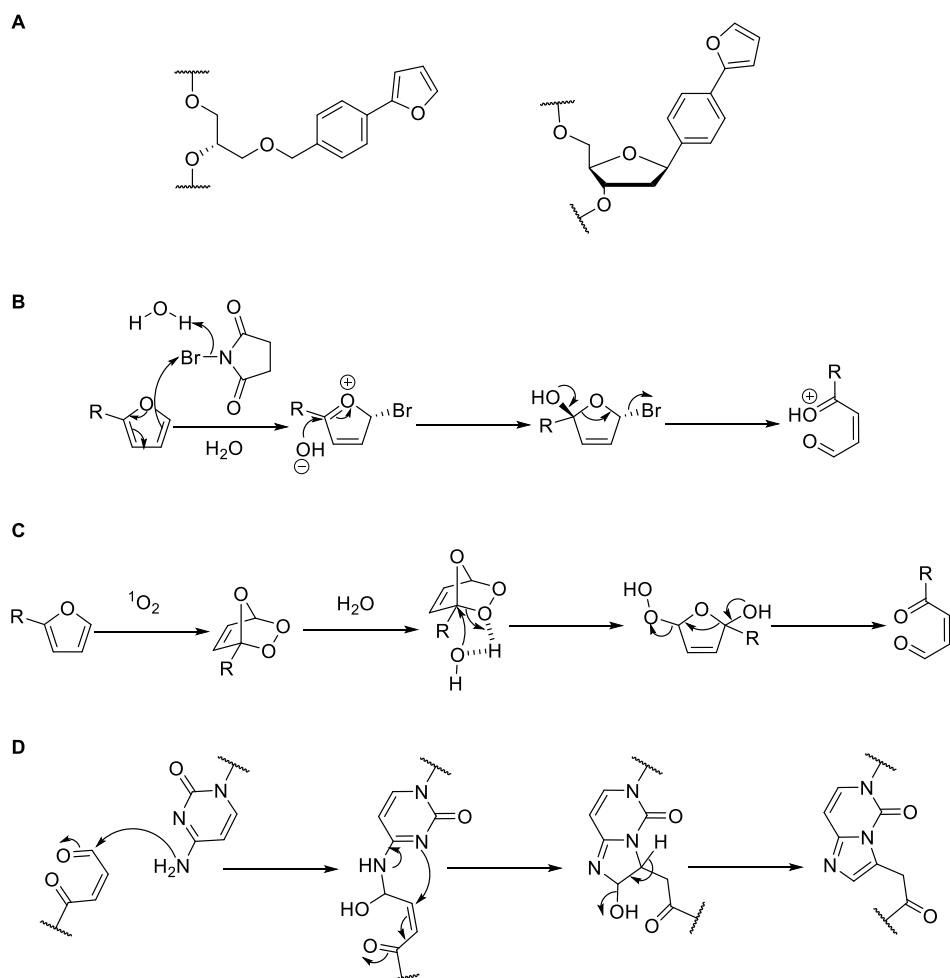
**Scheme 15** | DNA modification patterns with  $S_N1$  (A) and  $S_N2$  type alkylation reagents (B).

In addition to the direct alkylation with strong electrophiles there are several milder approaches, allowing a certain degree of selectivity of the modification. One such technique is the 3'-phosphate alkylation. It was first developed for RNAs cleaved in the presence of  $Mn^{2+}$  and imidazole.<sup>38</sup> Such cleavage leaves a 2',3'-cyclic phosphate at the 3'-end of the RNA chain (Scheme 16A). The phosphate is then reacted with a labeled arylmethyl halide to yield the 3'-phosphate tagged RNA. The method was later extended to the 3'-phosphate modification of both DNA and RNA by using aryl diazomethane.<sup>39</sup> Appropriate tailoring of the aryl moiety assures efficient introduction of the desired label at the nucleic acid 3'-end (Scheme 16B).



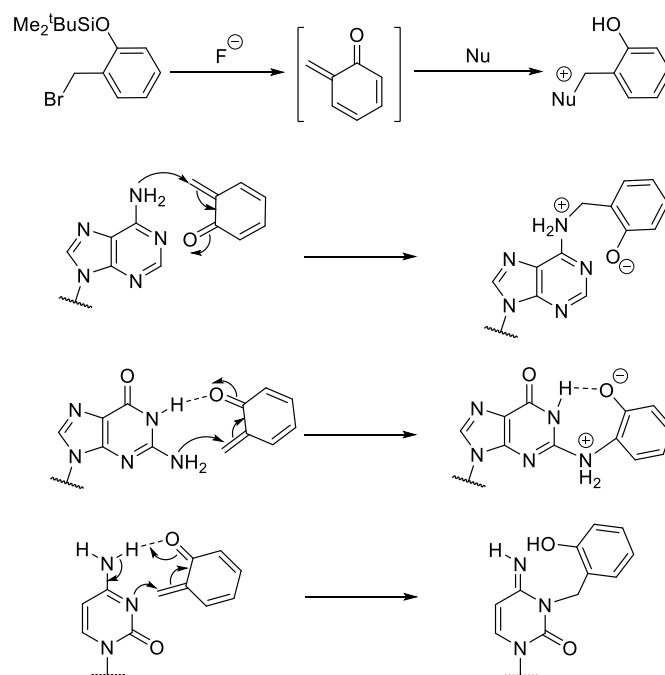
**Scheme 16** | 3'-End phosphate labeling: modification of the 2',3'-cyclic phosphate with an arylmethyl bromide (A),<sup>38</sup> and 3'-phosphate biotinylation with an aryldiazomethane derivative (B)

A group of techniques with a broad utility are the methods for DNA cross-linking. Usually the cross-linking probes consist of a complementary to the target DNA strand with an activatable functionality installed at a specific position. Hybridization of the two strands and activation of the probe delivers the modification to the target in a sequence-specific manner. One such technique is the DNA cross-linking by *in situ* furan oxidation demonstrated by the group of Madder.<sup>40,41,42,43</sup> The furan moiety is installed inside the modifying chain, either at C-2' on the backbone,<sup>40,41</sup> or through an abasic site for optimal hybridization with the target chain (Scheme 17A).<sup>42,43</sup> The furan oxidation can be carried out with *N*-bromosuccinimide (Scheme 17B),<sup>40,41,42</sup> or with singlet oxygen, generated with red light irradiation and methylene blue as a photosensitizer (Scheme 17C).<sup>43</sup> A plausible mechanistic model for the reaction of the 4-oxo-enal formed from the furan oxidation with the complementary cytosine base of the target chain affording a stable cross-link product is outlined in Scheme 17D.



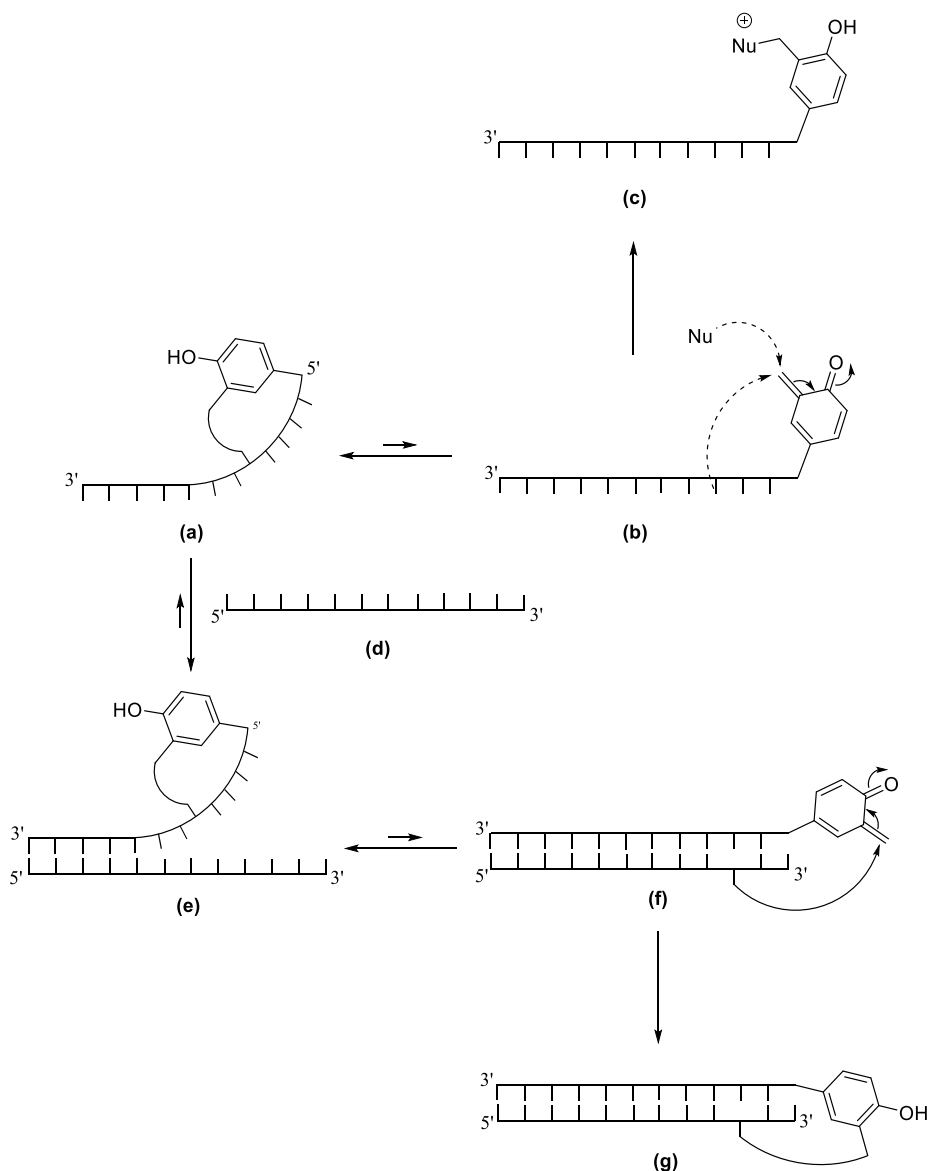
**Scheme 17** | Furan-oxidation triggered sequence-specific DNA cross-linking: optimized DNA monomer structures for introduction of the furan moiety (A), plausible mechanism of furan oxidation by *N*-bromosuccinimide (B) and singlet oxygen in aqueous medium (C), and cross-linking with a complementary cytosine base affording the stable cross-link product (D).

Another method for sequence-specific DNA cross-linking has been developed by the group of Rokita. Their initial studies have shown that chemical modification of simple deoxynucleosides and DNA can be achieved with *in situ* generated *o*-quinone methides.<sup>44</sup> The reactive species is obtained from an *O*-silyl protected *o*-bromomethylphenol in the presence of fluoride and is shown to react with nucleobases containing an exocyclic amino group - adenine, guanine and cytosine (Scheme 18).



**Scheme 18** | *o*-Quinone methide formation and possible reaction products with adenine, guanine and cytosine nucleobases

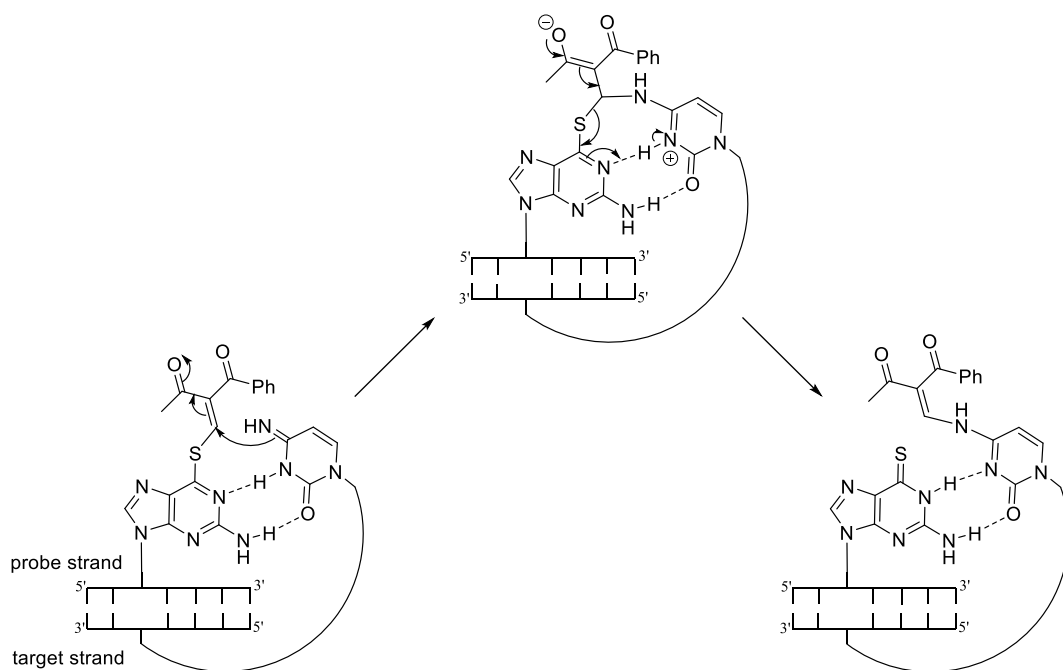
For cross-link applications<sup>45</sup> an analogous derivative was attached to the 5'-end of a carrier DNA, complementary to the target nucleic acid sequence. Upon treatment with fluoride the reactive quinone methide generated on the chain's end forms a self-adduct with one of the nucleobases on the modifying strand (Scheme 19, **a**). Such adduct formation is suggested to be reversible and is in equilibrium with activated species **b**. Species **b** can in turn be attacked by a nucleobase on the strand and form back **a** or by an external nucleophile Nu and form species **c**. Once the target chain **d** is added DNA duplex **e** forms, which is in equilibrium with quinone methide **f**. Species **f** can engage in an intrastrand interaction affording intermediate **e**, or form a cross-link adduct **g** with full base pairing. In practice the formation of such cross-link product seems to be favored. That is probably due to the energy gain from the additional base pairing, which is the energetic barrier preventing the back-formation of any constrained intermediates **e**.



**Scheme 19** | DNA cross-linking with an *on-strand* generated *o*-quinone methide – a mechanistic rationale for the exclusive formation of the interstrand product.

The principle of guided modification has been further extended to nucleic acid tailoring by functionality transfer developed by the group of Sasaki.<sup>46</sup> Their method for sequence-specific RNA modification relies on the use of a complementary DNA oligonucleotide probe, carrying an *S*-alkylated 6-thioguanosine base at the desired position for modification (Scheme 20). The *S*-alkyl moiety has a character of a Michael acceptor and is transferred to the complementary cytosine nucleobase of the target strand upon hybridization. The initial attack of the exocyclic amino group of cytosine on the 1,4-conjugate system appears to be triggered by the close proximity between the *S*-alkyl function and the target cytosine, brought by the DNA-RNA duplex formation. This strategy is unique in the sense that it not only delivers the modification sequence-specifically, but also allows for the release of the guiding strand without formation of cross-link products.



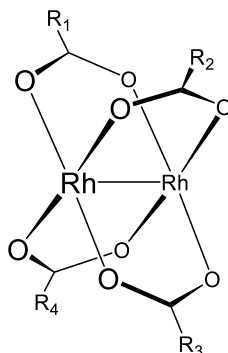


**Scheme 20** | Functionality transfer strategy developed by Sasaki et al.<sup>46</sup>

All post-synthetic approaches discussed so far except the direct alkylation with strong electrophiles insure the site-specific introduction of the desired functionality in a chemoselective manner and have proven valuable in a variety of specific tasks. They display, however, a few drawbacks that significantly narrow the scope of potential applications. For the methyltransferase approach that is mainly the variety of modified cofactors that the enzyme utilizes. One way to partially alleviate this problem is by re-designing the enzymes's cofactor binding pocket. Studies with the model MTase *M.HhaI* have shown that a triple replacement of nonessential residues at the cofactor binding site leads to an increased catalytic efficiency of the alkyl transfer of some sterically demanding moieties.<sup>47</sup> The chemical methods, on the other hand, although efficient and site-specific, are exclusively stoichiometric. A catalytic approach would be a viable alternative to the chemical and enzymatic methods described, as it combines increased group tolerance, thus greater substrate utility, with high reaction efficiency and chemoselectivity. Insertion reactions of metal-stabilized carbenoids are particularly well-suited for such purposes. They are efficient, require mild reaction conditions and have been applied for a broad variety of substrates. Two metals, rhodium(II) and copper(I), traditionally used in carbenoid chemistry, seem to be good candidates for such catalysts.

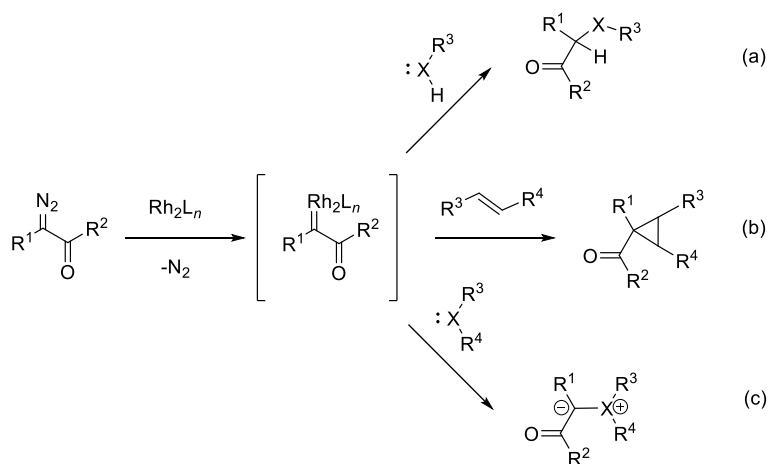
### 1.3. Rhodium(II)-catalyzed X-H insertion reactions

Dirhodium complexes are a unique class of compounds, containing a two-atom metal core ( $\text{Rh}^{\text{II}}\text{-Rh}^{\text{II}}$ ) with up to four bridging bidentate ligands in the equatorial positions, and two additional axial coordination sites (Scheme 21).



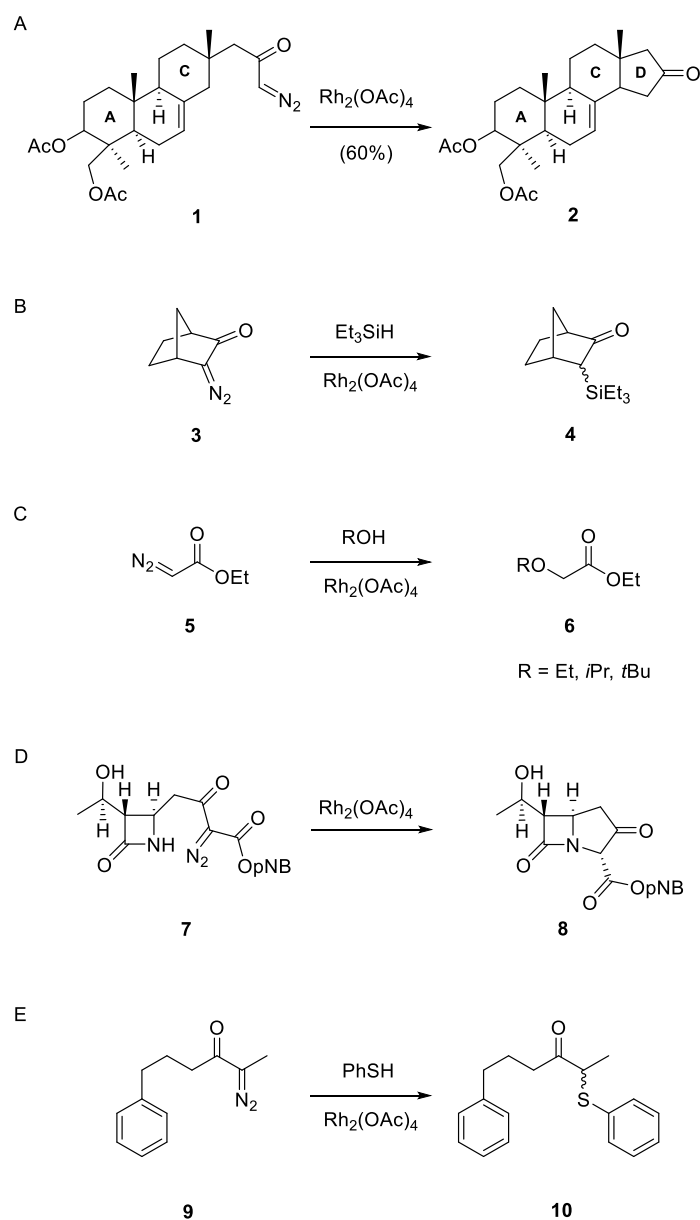
**Scheme 21** | General structure of a dirhodium tetracarboxylate.

Such dirhodium species display quite interesting coordination patterns and chemical reactivity. In particular, their reaction with  $\alpha$ -diazocarbonyl compounds gives rise to  $\text{Rh}(\text{II})$ -stabilized carbenoids. These carbenoids are electrophilic and depending on the nucleophile present they can undergo a variety of transformations, such as insertions into polar bonds, cyclopropanations, and ylide formation to afford a great diversity of products (Scheme 22).<sup>48</sup>



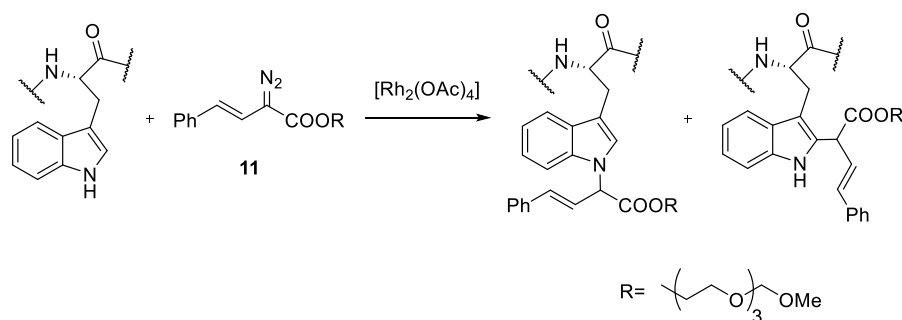
**Scheme 22** | Rhodium(II)-catalyzed X-H insertion (a), cyclopropanation (b), and ylide formation (c).

A class of reactions with a wide scope of practical applications are insertions into polar bonds of the type X-H, where X=C, Si, O, N, S. Scheme 23 shows several seminal examples for chemical transformations employing such X-H insertions reactions.



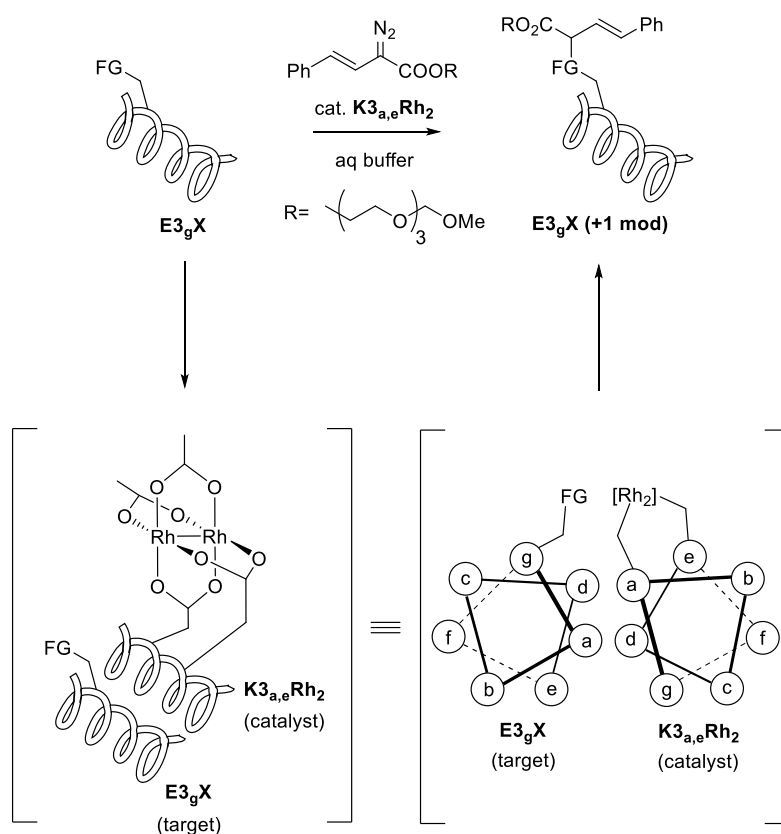
**Scheme 23** | Examples of X-H insertion reactions: a C-H carbenoid insertion closes up the D-ring on the steroid skeleton **2** (A);<sup>49</sup> a Si-H insertion for the preparation of the  $\alpha$ -silyl carbonyl adduct **4** (B);<sup>50</sup> an O-H insertion with ethyl diazoacetate was used for the preparation of alkyl ethers **6** (C);<sup>51</sup> an insertion into the N-H of the  $\beta$ -lactam **7** as the key step of the thienamycin synthesis proposed by Merck (D);<sup>52</sup> and a convenient route to unsymmetrically sulfenylated ketones **10** (E).<sup>53</sup>

The implementation of metal carbene chemistry for biomolecular modification is of particular interest for the present study. Protein alkylation with rhodium(II)-stabilized carbenoids in aqueous medium as a bioconjugation strategy has been first suggested by Francis et al.<sup>54,55</sup> They have demonstrated selective carbenoid insertion at N-1 and C-2 of the indole ring of exposed tryptophan residues (Scheme 24), accompanied by an excellent protein recovery and a moderate to good protein conversion.



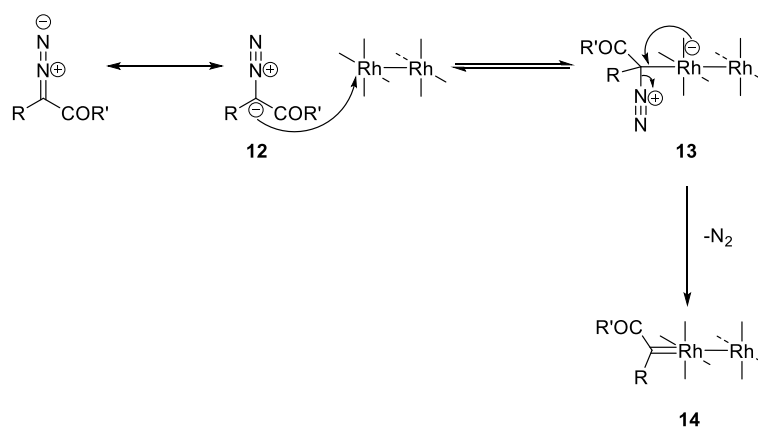
**Scheme 24** | Covalent modification of tryptophan residues in proteins with Rh(II)-carbenoids

Later on Ball and Popp have utilized the same approach but with a peptide-ligated rhodium catalyst for modification, guided by coiled-coil peptide assembly.<sup>56,57</sup> The specific binding of the catalyst-bearing peptide to the target delivers the metal carbenoid in the direct proximity of the modifiable functionality, resulting in a significant enhancement of its reactivity. This broadens the scope of accessible substrates, allowing for the efficient modification of a number of amino acid side-chains in addition to tryptophan (Scheme 25).



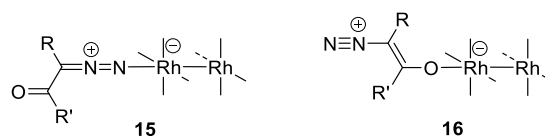
**Scheme 25** | Site-specific covalent modification of amino acid side-chains with Rh(II)-carbenoids: the metal carbenoid is delivered in close proximity to the modifiable group due to the coiled-coil interactions of the target and the catalyst chains.

On the other hand, it increases the selectivity for proximally located over randomly distributed potentially reactive groups, making selective targeting possible even in complicated mixtures, as demonstrated by the biotinylation of MBP-E3gW protein in total *E.coli* cell lysates.<sup>58</sup> The generally accepted mechanism of carbenoid formation is analogous to that proposed by Yates for the copper-catalyzed decomposition of  $\alpha$ -diazocarbonyl compounds (Scheme 26).<sup>59</sup> It involves an initial coordination of the partially negatively charged  $\alpha$ -carbonyl carbon of the diazo substrate **12** to one of the axial positions of the dirhodium species, followed by loss of nitrogen from the intermediate **13** to afford the metal-stabilized carbenoid **14**.



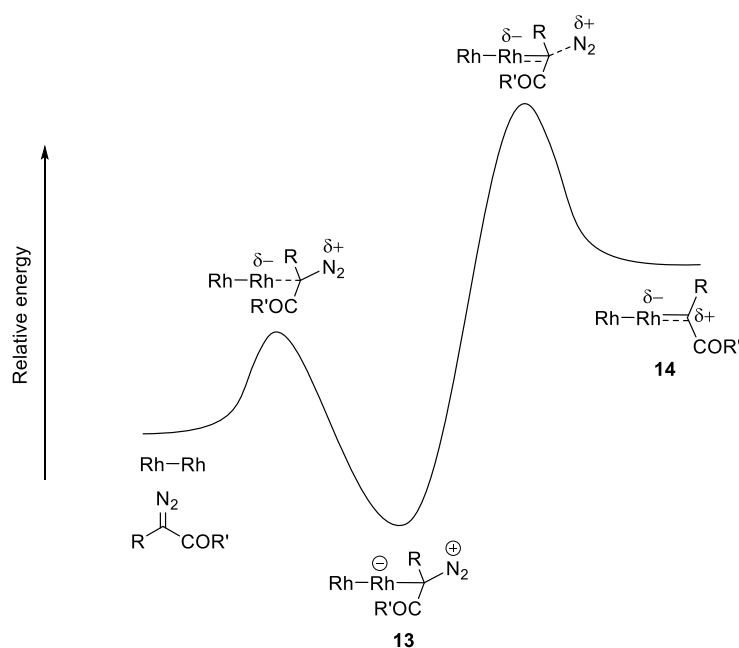
**Scheme 26** Generally accepted mechanism for the formation of rhodium-stabilized carbenes.

The formation of the rhodium species **14** as the most likely reaction pathway has been suggested as early as in the work of Teyssie et al. on the cyclopropanation of alkenes with alkyldiazoacetates in the presence of rhodium carboxylates.<sup>60</sup> Later on Pirrung and Morehead<sup>61,62</sup> proposed a saturation kinetics treatment of the dirhodium catalysis with the coordination of the diazo substrate being fast and reversible (pre-equilibrium), and the loss of nitrogen being the rate-limiting step. The observed mixed type inhibition pattern with a number of axial binders (aromatic hydrocarbons and Lewis bases) suggests formation of only one carbenoid per molecule of catalyst, as the catalytic efficiency of the second axial site decreases upon coordination at the first one. In addition to that, Davies and Singleton<sup>63</sup> show that in some cases the coordination of the  $\alpha$ -diazocarbonyl compound itself could proceed in an unproductive mode with the terminal nitrogen of the diazo group (**15**) or the carbonyl oxygen binding reversibly to the axial site of the rhodium catalyst (**16**) (Scheme 27). The electronics of the dirhodium core were further examined by Nakamura and co-workers<sup>64</sup> who demonstrated the role of the back-donation upon initial coordination of the diazo compound for facilitating the nitrogen extrusion and the carbenoid formation, and the enhancement of the carbene reactivity by the electron-withdrawing effect of the equatorial carboxylate ligands.



**Scheme 27** | Unproductive modes of coordination of the  $\alpha$ -diazo carbonyl compound to the dirhodium catalyst

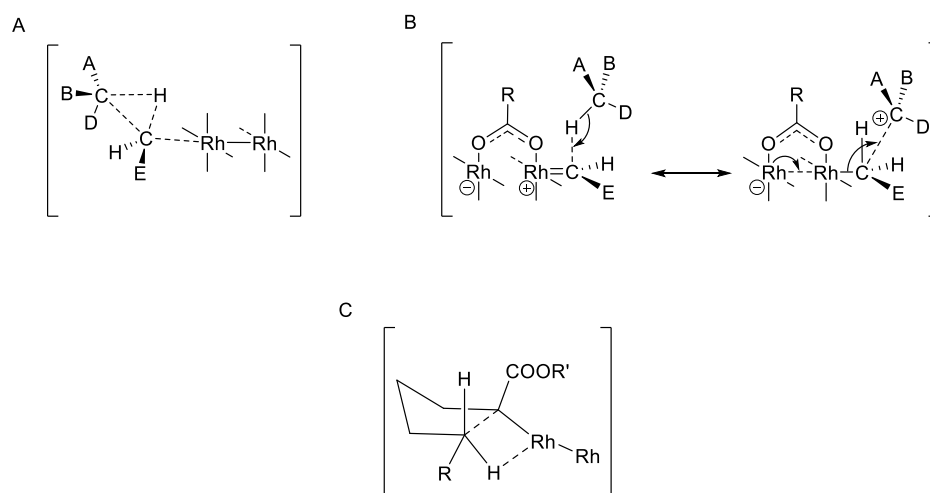
Several theoretical studies<sup>62,63,64</sup> have shown the free energy profile of the carbenoid formation step follows the general trend shown on Scheme 28. Coordination of diazo compound **12** to the dirhodium species results in a complex lower in energy (**13**) which then overcomes a relatively high activation energy barrier to expel nitrogen and afford the rhodium-stabilized carbenoid **14**.



**Scheme 28** | General trend of the free energy profile of rhodium-catalyzed carbenoid formation

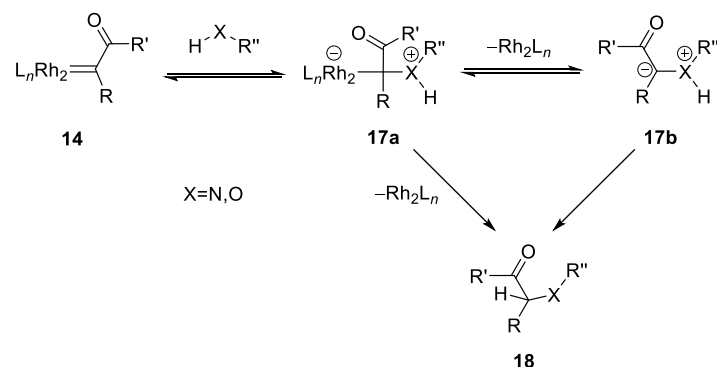
Concerning the mechanism of the insertion step several different cases should be considered depending on the polarity of the X-H bond. Direct kinetic investigation of the insertion process has proven difficult since in almost all cases the loss of nitrogen is the rate limiting step.<sup>64</sup> A widely accepted mechanism for the reaction of C-H insertion involves the formation of a concerted three-center transition state (Scheme 29A) as proposed by Doyle et al.<sup>65</sup> for a case of intramolecular cyclization in substituted  $\alpha$ -diazo esters. In a later study Nakamura et al.<sup>64</sup> show the C-H insertion transition state is concerted but highly asynchronous with two events taking place as demonstrated by the corresponding resonance structures: a hydride transfer from the alkane to the carbene, and C-C bond formation and release of the ligated metal (Scheme 29B). A different mechanism has been proposed by the group of Taber.<sup>66</sup> They hypothesize that the interaction between the electron-deficient carbene carbon and the target C-H in the form of a three-centered species is rapid and reversible, and

does not lead to the formation of the desired product. Instead, the reaction takes place through electron density redistribution in a different, four-centered transition state in which the target C-H is aligned with the Rh-C bond (Scheme 29C). A concerted C-H activation/C-C bond formation is consistent with retention of the configuration at the target C-H.<sup>67</sup> A similar reaction pathway has been proposed for the reaction of Si-H insertion.<sup>68</sup>



**Scheme 29** | Concerted transition states for the rhodium-catalyzed C-H insertion: according to Doyle (A),<sup>65</sup> Nakamura (B),<sup>64</sup> and Taber (C)<sup>66</sup>

The mechanism of carbene insertion into polar bonds, such as N-H and O-H, is much less clear. There exist two main contradicting models: a concerted transition state very much alike the one observed with the C-H insertion, and a step-wise mechanism with formation of an ylide intermediate **17a** or **b**, followed by a 1,2-proton shift (Scheme 30).<sup>69</sup> Discussions arguing a concerted transition state being the more likely reaction pathway are quite inconclusive and solely based on indirect kinetic data<sup>70</sup> or stereochemical reaction outcomes.<sup>71</sup> A good rationale for the latter mechanism are the experiments of Hu<sup>72</sup> and Davies,<sup>73</sup> in which an oxonium ylide, generated *in situ* from a rhodium carbenoid, is trapped with an imine, respectively subjected to a [2,3]-sigmatropic rearrangement. Another point, which should also be considered, is that the 1,2-proton shift could take place in the metal-bound (**17a**) or in the metal-free ylide (**17b**).

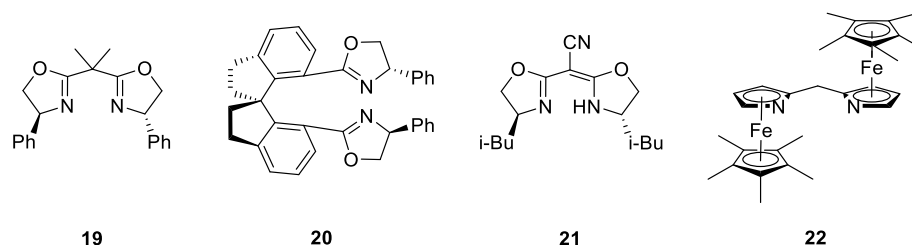


**Scheme 30** | Step-wise mechanism of X-H insertion in polar bonds (X=N,O) involving the formation of an ylide (**17a** and **b**)

#### 1.4. Copper(I)-catalyzed X-H insertion reactions

X-H insertion reactions with copper(I)-stabilized carbenoids derived from  $\alpha$ -diazo carbonyl compounds have been known for decades, yet their synthetic significance and potential application have been overshadowed by other better developed catalytic systems. The works of Jørgensen,<sup>74</sup> Fu<sup>75,76</sup> and Zhou<sup>77</sup> have refocused the attention back to copper, demonstrating its catalytic efficacy in a number of stereoselective transformations of practical significance.

The copper catalyst for such insertion reactions is usually formed *in situ* by mixing a catalyst precursor (e.g.  $Cu(OTf)_2$ ,  $CuBr$ ,  $CuCl$ ,  $Cu(MeCN)_4PF_6$ ) with a strongly coordinating chiral ligand, such as bis(oxazoline) (**19**) and (**20**), semicorrin (**21**), or bis(azaferrocene) (**22**) (Scheme 31).<sup>74,75,76,77</sup>



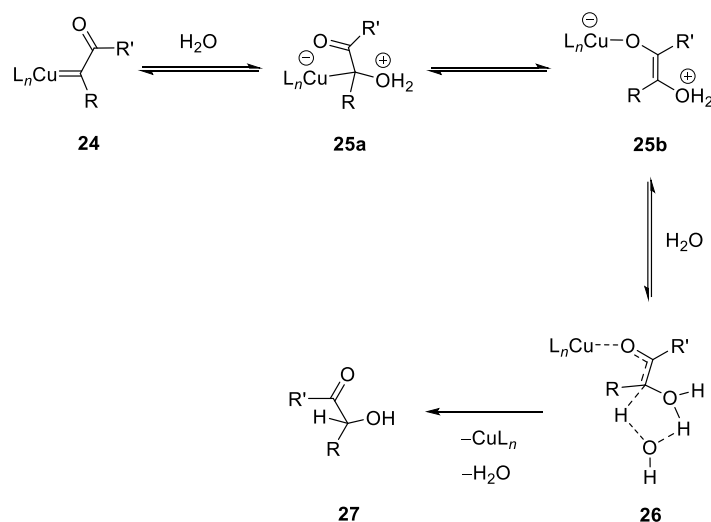
**Scheme 31** | Examples of copper(I) ligands for asymmetric copper(I)-catalyzed N-H and O-H insertion reactions

Concerning the oxidation state of the metal both Kochi<sup>78</sup> and Teyssie<sup>60</sup> have shown that in the presence of an  $\alpha$ -diazo carbonyl compound copper(II) is reduced to copper(I), which is the catalytically active species.

Mechanistically the copper(I)-catalyzed polar X-H insertions (X=N, O) follow the general layout proposed for rhodium(II), involving coordination of the  $\alpha$ -diazo carbonyl compound to the metal catalyst, loss of nitrogen, and insertion of the electrophilic carbene formed into the X-H bond *via* the formation of an ylide intermediate and a subsequent 1,2-proton shift. In a computational study of the O-H insertion in water with a copper(I)/bis(oxazoline) catalyst Yu et al.<sup>79</sup> suggest the insertion step

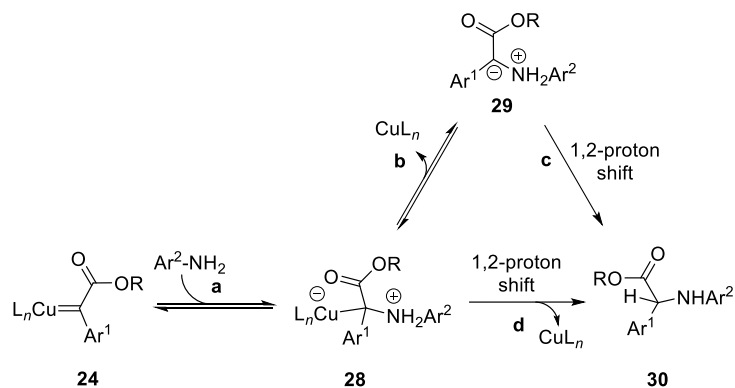


includes migration of the copper species from the  $\alpha$ -carbonyl carbon to the carbonyl oxygen with the formation of enolate **25b** and a subsequent water-assisted proton transfer (**26**) with the metal catalyst tightly associated to the ylide intermediate (Scheme 32).



**Scheme 32** | Reaction mechanism for the copper(I)-catalyzed carbenoid O-H insertion in water as proposed by Yu et al.<sup>79</sup>

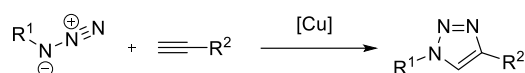
In a later study on N-H insertions with a well-defined binuclear chiral copper(I) catalyst Zhou et al.<sup>77</sup> demonstrate that the affinity of the metal catalyst towards the ylide can be tuned by changing the electronic characteristics of the substrates making both the metal-free (b→c) and the metal-associated pathways (d) possible (Scheme 33). They've also shown the 1,2-proton transfer is the rate-limiting step of the insertion reaction, which is consistent with the earlier observations of the group of Fu.<sup>75,76</sup>



**Scheme 33** | Possible reaction pathways for the copper(I)-catalyzed insertion of an aryl diazo alkyl ester into the N-H of an aniline as devised by Zhou et al.<sup>77</sup>

## 1.5. Copper(I)-catalyzed azide-alkyne cycloaddition

In addition to being suitable for carbenoid stabilization, copper(I) is the most widely used catalyst for [2+3] azide-alkyne cycloaddition, or copper ‘click’ chemistry – a reaction of broad utility for chemical biology and biochemistry. The CuAAC, independently reported by Fokin and Sharpless,<sup>80</sup> and by Meldal,<sup>81</sup> (Scheme 34) is a star example for a ‘click’ reaction, as defined by Sharpless, a transformation easy to perform, resulting in high yields, with little or no by-products, carried out under mild conditions and unaffected by the nature of the functional groups on the entities being connected.<sup>82</sup> In this process an azide and a terminal alkyne selectively form a 1,4-disubstituted 1,2,3-triazole ring in a reaction of copper(I)-catalyzed cycloaddition.



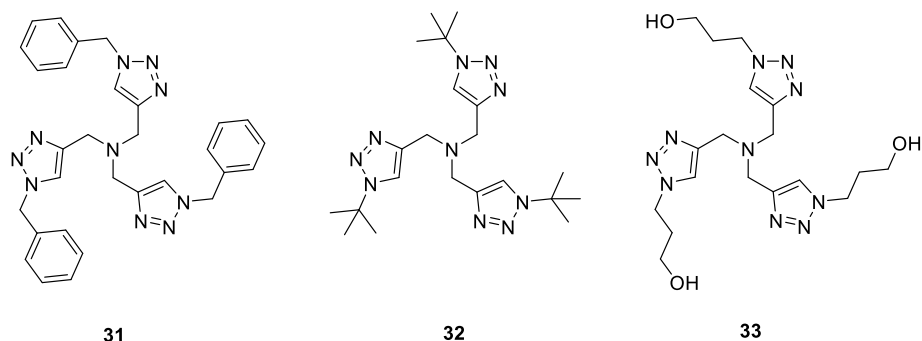
**Scheme 34** | Copper(I)-catalyzed azide-alkyne cycloaddition.

The unique combination of several factors makes this transformation the coupling reaction of choice, especially when it comes to applications for bioconjugation.<sup>83,84</sup> The two starting components alone are chemically relatively inert, and their reactivity remains almost unaffected by the steric and electronic properties of the groups attached.<sup>33</sup> The reaction takes place in aprotic and protic solvents, and tolerates a large variety of functional groups. The 1,2,3-triazole ring formed is stable to hydrolysis, oxidation and reduction, has a large dipole moment (approximately equal to that of an amide bond), aromatic character and is able to participate in hydrogen bonding.<sup>84,85</sup>

Copper in all three oxidation states (0, +I, +II) can be used as a metal source.<sup>85</sup> Copper(I) salts (iodide, bromide, chloride, acetate) and complexes (e.g.  $[\text{Cu}(\text{CH}_3\text{CN})_4]\text{PF}_6$ ,  $[\text{Cu}(\text{CH}_3\text{CN})_4]\text{OTf}$ ), as well as Cu(0), which is oxidized to the catalytically active Cu(I) by the molecular oxygen present in the reaction mixture, are all viable options. When using a copper(I) source oxygen-free conditions are required, as copper(I) gets oxidized to Cu(II) or disproportionates to Cu(0) and Cu(II). The use of Cu(II) salts is accompanied by sacrificial oxidation of the alkyne substrate to generate Cu(I) species, required for the catalysis. The combination of a copper(II) salt, such as copper(II)sulfate or copper(II)acetate and a mild reductant like ascorbate turns out to be a convenient strategy for the *in situ* generation of the copper(I) catalyst, particularly in aqueous systems used in bioconjugation reactions.<sup>80,86</sup> The successful implementation of such a water-based copper(I) catalytic system is complemented by the use of appropriate ligands.

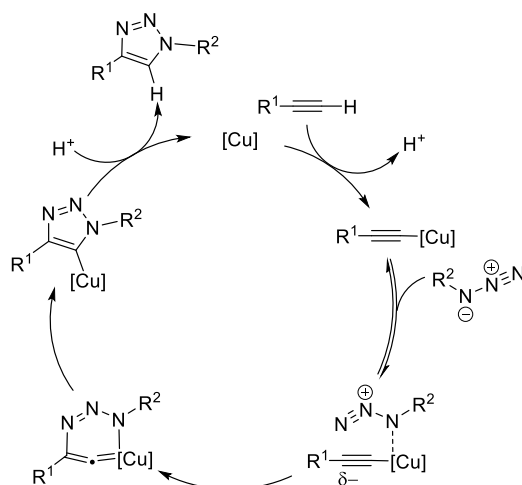
A group of triazole-based ligands has gained popularity for such aqueous systems for CuAAC. Among them, tris(benzyltriazolylmethyl) amine (TBTA, **31**),<sup>87</sup> tris(tert-butyltriazolylmethyl)amine (TTTA, **32**) and tris(3-hydroxypropyltriazolylmethyl)amine (THPTA, **33**)<sup>86</sup> (Scheme 35) are of particular

importance due to their ability to stabilize the copper(I) species, accelerate the cycloaddition reaction and scavenge the generated reactive oxygen species, which could otherwise cause oxidative damage to the target biomolecules.



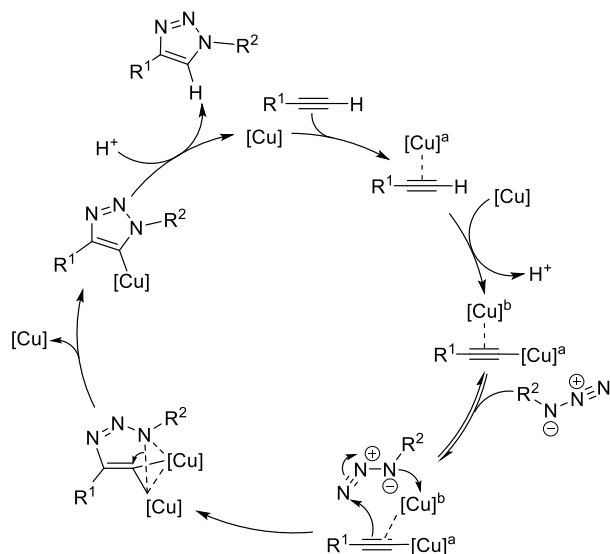
**Scheme 35** | Triazole-based ligands for CuAAC in water.

A widely accepted mechanism for the CuAAC is the one depicted in Scheme 36.<sup>80</sup> The formation of a copper acetylide is followed by coordination of the quite nucleophilic N-3 of the azide to the metal center. Back-donation from the metal enhances the nucleophilicity of the  $\beta$ -acetylide carbon, which attacks the terminal electrophilic N-1 of the azide, forming a six-membered cyclic intermediate, which then rearranges to a five-membered copper triazolide. Protonation of C-5 releases the triazole product and regenerates the metal catalyst. In a recent study the group of Fokin stipulates the involvement of not one, but two copper atoms in the catalytic process.<sup>88</sup> They have shown that monomeric copper acetylides are unreactive towards azides, unless an exogenous copper catalyst is added. Crossover experiments with isotopically enriched exogenous copper species have led them to devise the mechanism shown on Scheme 37.



**Scheme 36** | Mechanism of the CuAAC as initially suggested by Fokin and Sharpless.

One copper atom takes part exclusively in  $\pi$ -complexation with the acetylene system, a second one acts purely as a  $\sigma$ -bond ligand, forming a copper acetylide. Reversible coordination of the azide is followed by a nucleophilic attack by the  $\beta$ -acetylide carbon. At this stage the ligand exchange on the two copper centers takes place faster than the next C-N bond forming event, allowing isotopic enrichment of the bimetal species. Once the C-N bond is formed, one metal atom leaves the system, yielding a five-membered copper triazolide, which after protonation of C-5 releases the triazole product and the second copper atom.



**Scheme 37** | Proposed catalytic model for CuAAC with two copper atoms.

## Chapter 2. Nucleic acid alkylation with Rh(II) carbenoids

The aim of the work discussed in this chapter was to establish a direct catalytic approach for post-synthetic DNA and RNA alkylation using rhodium(II)-stabilized carbenoids.

### 2.1. Alkylation of single-stranded DNAs and RNAs

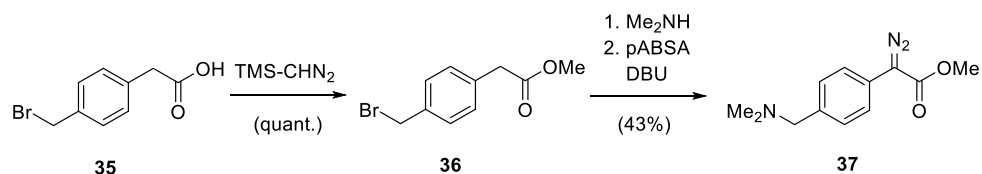
The rhodium(II)-catalyzed conversion of  $\alpha$ -diazo carbonyl compounds has been established as a powerful approach for generating rhodium carbenoids. These compounds could then undergo a number of useful transformations such as insertions into polar bonds. Rhodium(II) carboxylates are also known for their ability to bind DNA and RNA molecules.<sup>89</sup> The extensive structural characterization of the Rh(II)-DNA complexes has shown that such binding can occur in different ways depending on the oligonucleotide structure. This can lead to striking changes in DNA conformation that, in turn, can alter its physiological activity in the living cell.<sup>90,91,92</sup> On the other hand DNA and RNA possess a number of nucleophilic groups which can be involved in reactions with the rhodium(II)-stabilized carbenes. However, in the literature so far there is no precedence for addressing such a possibility. In addition, the interactions between the catalytic dirhodium species and the oligonucleotides could be exploited for the proper targeting and modulation of the activity of the catalyst.

The  $\alpha$ -diazo carbonyl substrate plays a decisive role in establishing an efficient metal-carbene system for nucleic acid alkylation. The diazo substrate **11**, commonly utilized for the amino acid side-chain modification in aqueous media,<sup>54,55,56,57</sup> is sparingly soluble in water and requires the use of a co-solvent. In addition, even when stored neat at  $-20^{\circ}\text{C}$ , it tends to form a cyclic side-product (**34**) most likely through a reaction of an intramolecular dipolar cycloaddition. To avoid these problems a re-design of the diazo substrate was done.



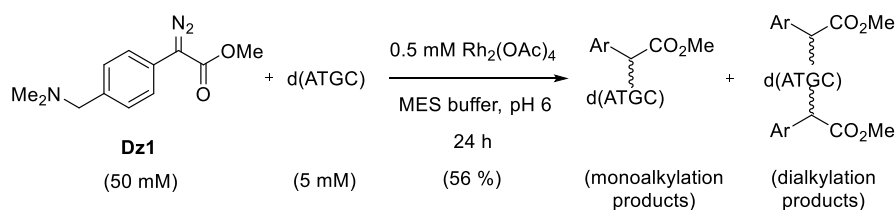
In the new  $\alpha$ -diazo substrate Dz1 (**37**) the  $\alpha$ -carbon was directly connected to the phenyl ring, omitting the 3,4-double bond fragment, the PEG ester group was replaced with a methyl group, and a *N,N*-dimethylaminomethyl residue was installed in *para*-position of the phenyl ring to improve the water solubility. Dz1 was accessed in only three synthetic steps: methylation of commercially available (4-bromomethyl)phenylacetic acid (**35**) under mild conditions, followed by a substitution of

the benzylic bromide and a nucleophilic diazo transfer afforded the target molecule in a moderate yield (Scheme 38). Dz1 was well-soluble in aqueous buffers at pH  $\leq 6$  (concentrations of up to 250mM) and was stable for several days at that pH at room temperature.



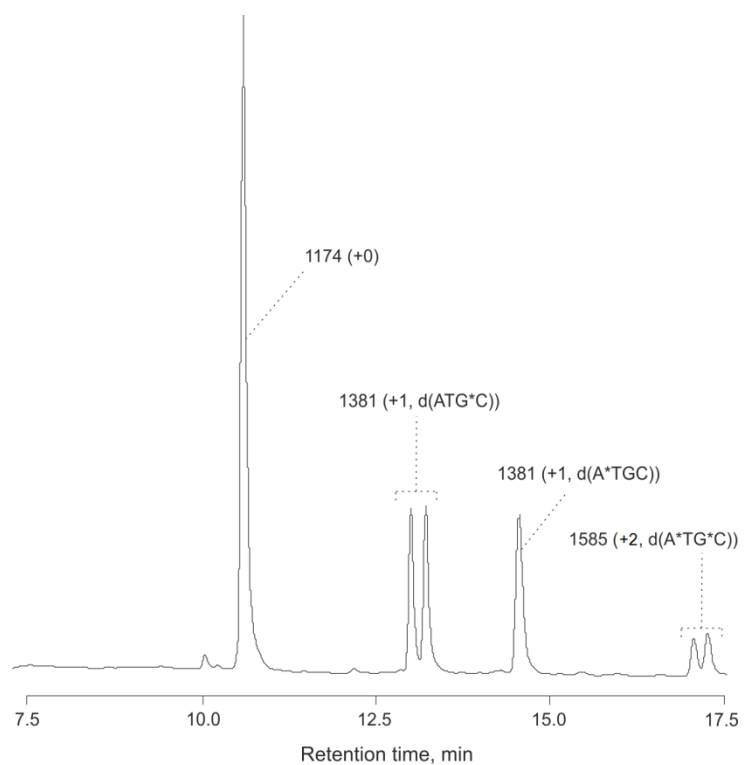
**Scheme 38** | Synthetic sequence for preparation of Dz1 (37).

In a proof-of-concept experiment a short single-stranded oligonucleotide d(ATGC), containing the four canonical DNA bases, was subjected to alkylation with Dz1 in the presence of 10 mol% Rh<sub>2</sub>(OAc)<sub>4</sub> in a buffered aqueous medium at room temperature (Scheme 39).

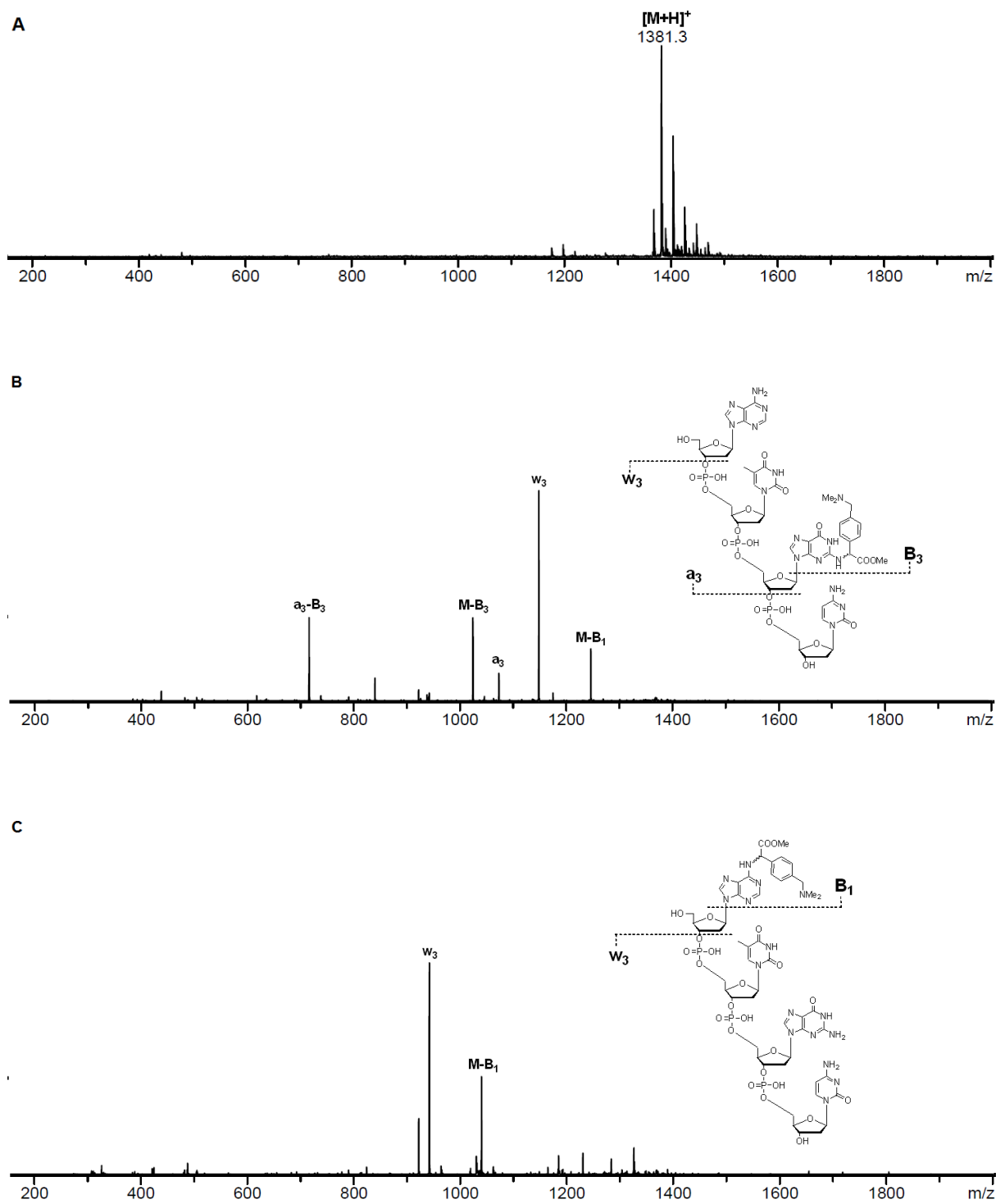


**Scheme 39** | Proof-of-concept experiment for rhodium-carbenoid alkylation.

HPLC/MS analysis of the reaction mixture after 24 h revealed that more than half of the starting oligonucleotide (56 %) had been converted to a number of singly and doubly modified oligonucleotide species (Figure 1). MS/MS analysis revealed that the modification had taken place exclusively on the purine nucleobases (Figure 2 and 3). A direct evidence for that were the molecular ion fragments after loss of the alkylated base (M-B)<sup>+</sup>, which were found in all cases. In addition to that the samples displayed the typical for DNA a/w-fragmentation patterns, with the expected a- and w-ions present.

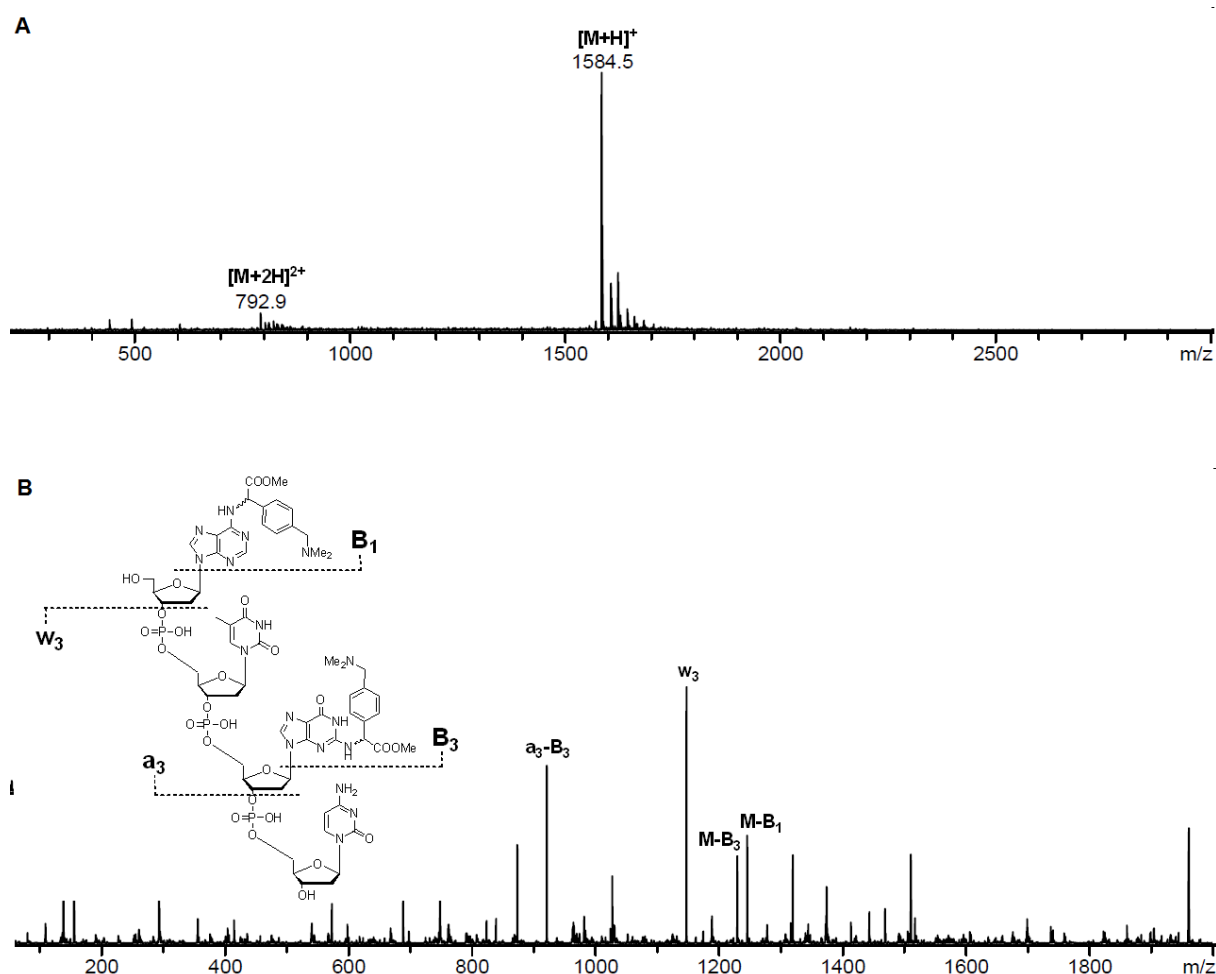


**Figure 1** | Alkylation of d(ATGC) with rhodium carbenoids: HPLC trace of the reaction mixture after 24 h. The numbers indicate the molecular weight of the oligonucleotide in the respective fraction. The number of alkylations per molecule of oligonucleotide and the modified base(\*) as determined by MS/MS are given in parenthesis.



**Figure 2** | Alkylation of d(ATGC) with rhodium carbenoids: ESI-MS (A) and MS/MS (B, C) characterization of fractions, containing monoalkylated products. The fragmentation patterns are consistent with a modification on guanine (B) and adenine (C).





**Figure 3** | Alkylation of d(ATGC) with rhodium carbenoids: ESI-MS (A) and MS/MS (B) characterization of a fraction, containing dialkylated products.

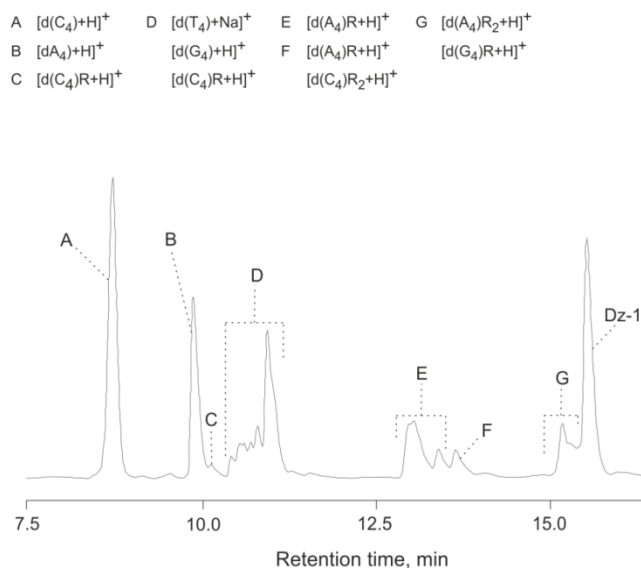
Experiments with the four homotetramers d(A)<sub>4</sub>, d(C)<sub>4</sub>, d(T)<sub>4</sub>, and r(U)<sub>4</sub> (Table 1, entries 1-4) were particularly important for assessing of the individual propensity of each base towards alkylation. The oligonucleotides d(C)<sub>4</sub> and d(A)<sub>4</sub> demonstrated good reactivity and were converted to a number of singly- and doubly-modified species at moderate conversions of the starting oligonucleotides. In contrast, the thymine and the uracil tetramers proved to be completely unreactive. This on one hand suggests that the alkylation does not take place on the ribo-phosphate backbone of the oligonucleotide, and on the other, that the functionality responsible for the cytosine and adenine reactivity is not present in the thymine and uracil nucleobases. It should be noted that the cytosine tetramer was efficiently alkylated in contrast to the data from the d(ATGC) experiment. This was accompanied by a complete conversion of the Dz1 in less than 3 h, which was not the case for d(A)<sub>4</sub>, d(T)<sub>4</sub>, and r(U)<sub>4</sub>, where only partial consumption of the diazo substrate was observed even after 24 h incubation, indicating decreased overall activity of the catalyst in the latter cases.

**Table 1** | Scope of the rhodium(II)-carbenoid alkylation with single-stranded DNA and RNA oligonucleotides.

Entry	Substrate	Number of alkylation products		Conversion of oligonucleotide, % <sup>[a]</sup>	Conversion of Dz1, % <sup>[a]</sup>	N-H/O-H selectivity factor <sup>[b]</sup>
		Monoalkylation	Dialkylation			
1	d(T) <sub>4</sub>	0	0	0	20	≈0
2	r(U) <sub>4</sub>	0	0	0	30	≈0
3	d(A) <sub>4</sub>	4	2	39	40	-
4 <sup>[c]</sup>	d(C) <sub>4</sub>	6	6	45	>98	-
5 <sup>[d]</sup>	d(TAT)	2	0	20	>98	282
6 <sup>[d,e]</sup>	d(TGT)	2	0	25	>98	282
7 <sup>[c,d]</sup>	d(TCT)	2	0	16	>98	210
8 <sup>[d]</sup>	d(ATGC)	4	2	56	>98	-
9	r(ACU GCU)	3	0	68	>98	9970
10	d(ACT GCT)	5	1	56	>98	-
11	d(CTC TCT)	2	0	32	>98	1550
12	d(CTG GCT)	6	0	42	>98	1050
13	d(TTT ATT TGT TTC TTT)	4	2	37	>98	-

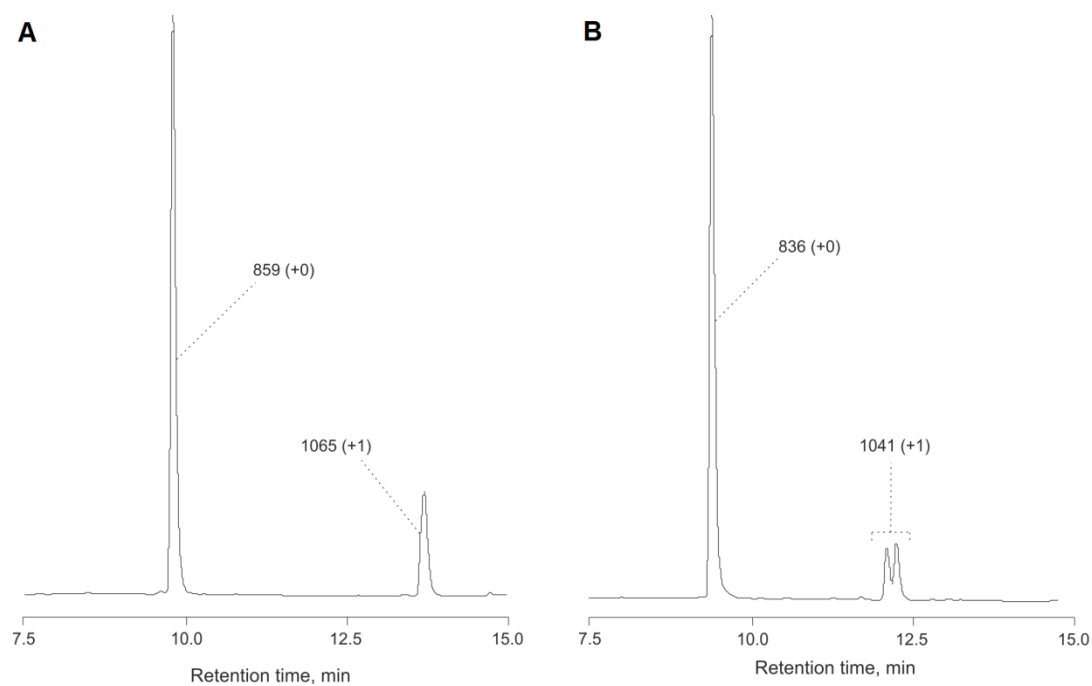
Reaction conditions: 5 mM oligonucleotide, 50 mM Dz1, 0.5 mM Rh<sub>2</sub>(OAc)<sub>4</sub>, 50 mM MES buffer, pH 6.0, 24 h at room temperature. [a] Conversion of the starting material, oligonucleotide or Dz1, determined by HPLC. [b] Determined by multiplying the ratio N-H/O-H insertion products by the effective concentration of N-H and O-H nucleophiles; calculated for cases with only monoalkylation products. [c] Dz1 was completely consumed after 3 h. [d] Alkylation sites determined by MS/MS analysis. [e] 5 mM d(TAT), 5 mM d(TGT), 1 mM Rh<sub>2</sub>(OAc)<sub>4</sub>.

The tetramer d(G)<sub>4</sub> displayed quite an interesting behavior, as attempts to alkylated it alone were unsuccessful. In a mixed experiment, however, in the presence of the other three DNA tetramers, it was efficiently modified (Figure 4). d(T)<sub>4</sub> proved again to be unreactive, as no traces of modified products were found.

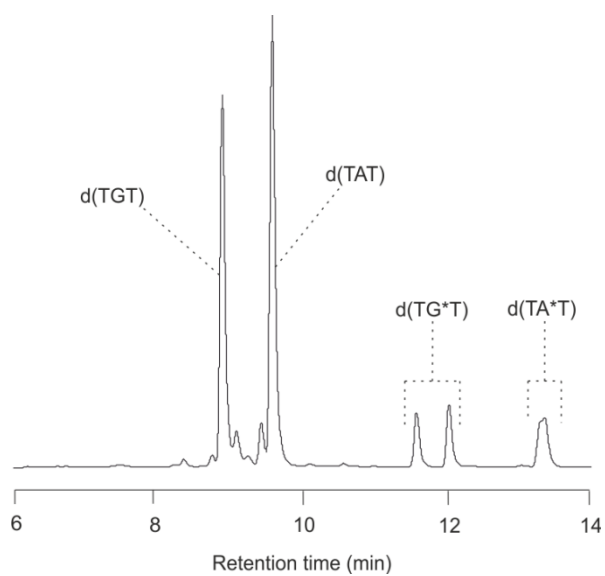


**Figure 4** | HPLC trace of the modification reaction of a mixture of  $d(A)_4$ ,  $d(C)_4$ ,  $d(G)_4$ , and  $d(T)_4$  with rhodium carbenoids. Peak identification was done by ESI-MS.

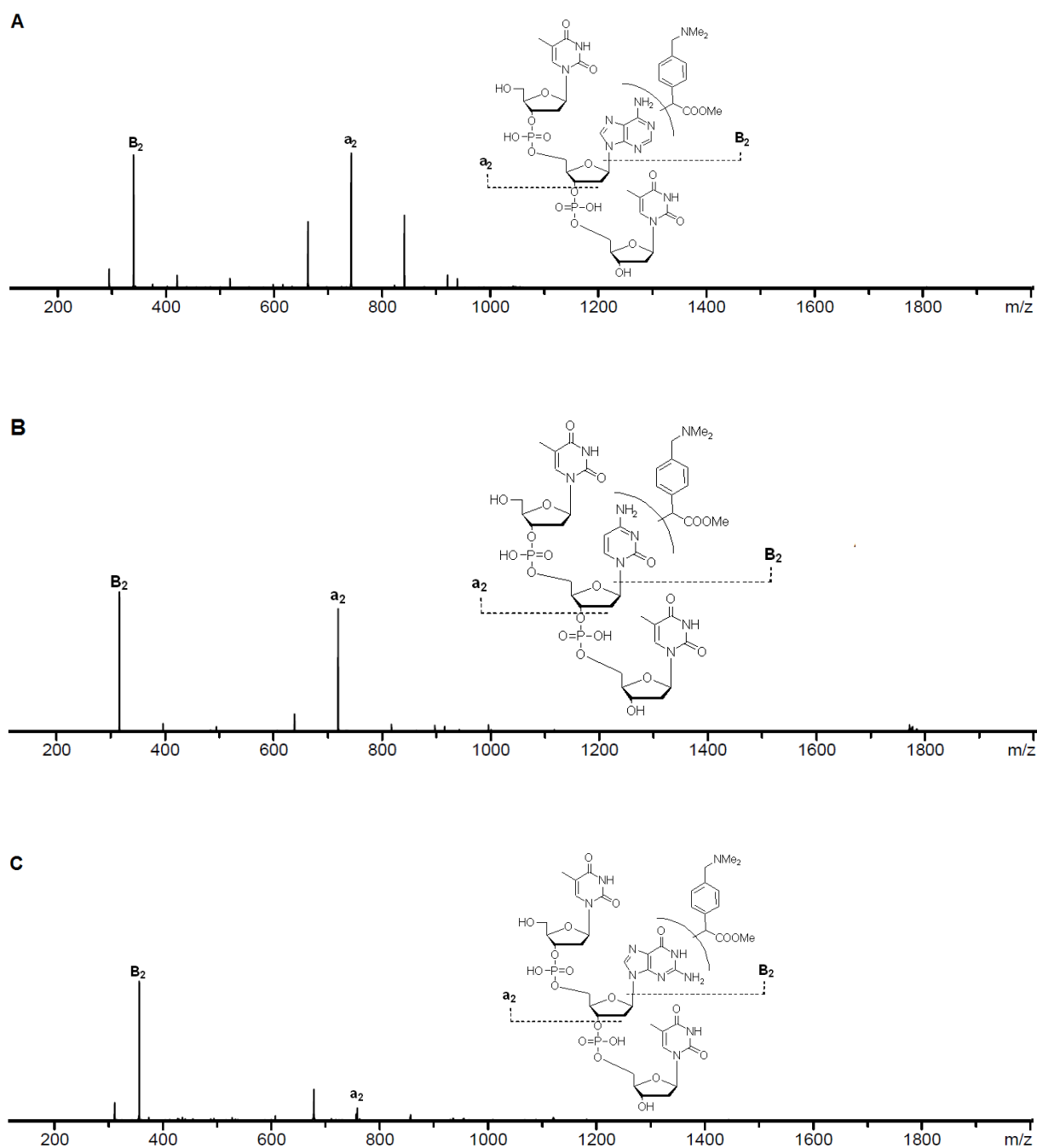
The modification experiments were continued on three nested oligonucleotide trimers  $d(TCT)$ ,  $d(TAT)$ , and  $d(TGT)$  in order to examine the individual reactivity of each nucleobase, assuming that no reaction took place on thymine or the ribo-phosphate backbone of the oligonucleotide. Once again alkylation of  $d(TGT)$  was only possible upon addition of a second oligonucleotide, in this case  $d(TAT)$ . In all three cases two monoalkylation products were formed as indicated by the HPLC/MS analysis of the reaction mixtures (Figure 5 and 6). The trimer  $d(TGT)$  showed the highest propensity for alkylation and  $d(TCT)$  the fastest consumption of Dz1. Tandem MS analysis of the modified products showed an intense signal for the modified  $B_2$ , accompanied by the respective  $a_2$ -ion - a clear evidence that the modification had taken place on the adenine, cytosine or guanine without affecting the thymine bases, the 2'-deoxyribose or the phosphate groups (Figure 7).



**Figure 5** | HPLC trace of the modification reaction of d(TAT) (A) and d(TCT) with Dz1. The product masses and the number of alkylations per oligonucleotide molecule (in parenthesis) are indicated for each peak.



**Figure 6** | HPLC trace of the modification reaction of a mixture of d(TGT) and d(TAT) with Dz1. Only monoalkylation products were detected. The asterisk after the base shows the position of modification established by MS/MS.



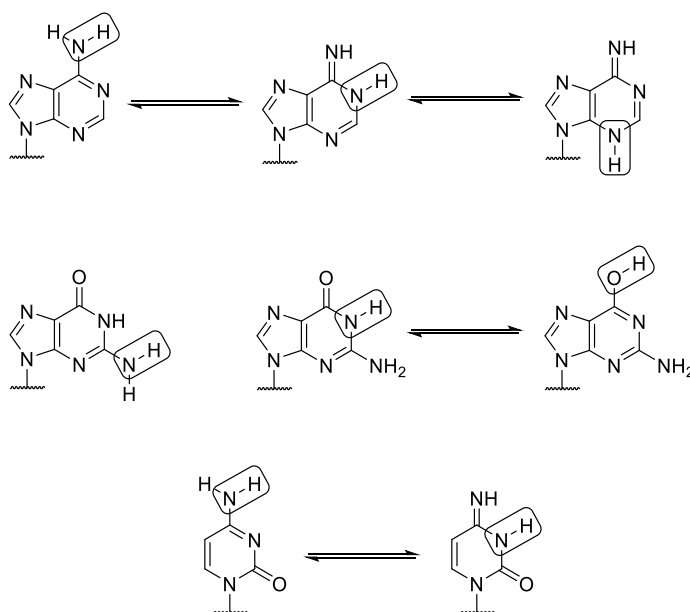
**Figure 7** | MS/MS fragmentation of singly-modified d(TAT) (A), d(TCT) (B), and d(TGT) (C) with Dz1.

The oligonucleotide substrate study was complemented by a few medium-sized DNA and RNA oligonucleotides, tested under the alkylation conditions described above (Table 1). In all cases (entries 9-13) moderate to good conversions were observed, as the presence of guanine in the structure generally increased the yield of modified oligonucleotide. Concerning the type of nucleic acid entries 9 and 10 demonstrate, that RNAs can not only be targeted, but in fact more efficiently alkylated than comparable DNAs.

If not used for nucleic acid alkylation, the diazo substrate was directed to O-H insertion in water, giving the secondary alcohol product. Assuming that the alkylation takes place on a nitrogen nucleophile on the base (N-H insertion) an N-H/O-H insertion selectivity factor was calculated for the examples which only give rise to monoalkylation products (Table 1). It shows a remarkable preference for N-H over O-H insertion despite of the overwhelming excess of water in the reaction system.

## 2.2 Regioselectivity of the alkylation. Determination of the alkylation sites: tandem MS and NMR experiments

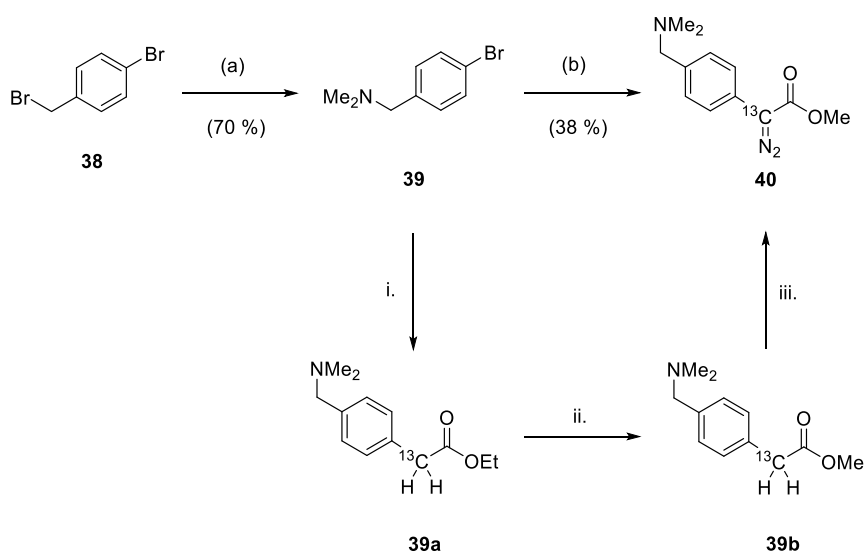
The reactivity of the homotetramers together with the information from the tandem MS experiments with d(ATGC) and the three nested trimers delivered a direct evidence for the involvement of the adenine, guanine and cytosine nucleobases in the alkylation process. Each nucleobase itself contains a number of reactive nucleophilic centers, which could take part in an insertion reaction with the rhodium carbenoid (Scheme 40).



**Scheme 40** | Potential reactive sites for insertion of the rhodium carbenoid.

A more rigorous structure characterization of the reaction products was needed in order to precisely define the alkylation site on the nucleobase and establish the regioselectivity of the modification reaction. Such structure elucidation was completed by NMR spectroscopy. In principle, the unambiguous assignment of NMR structures could be quite challenging even of short oligonucleotides due to the large number of similar proton and carbon resonances. In order to alleviate that problem a  $^{13}\text{C}$ -labeled analogue of Dz1 (**40**) was prepared. The  $^{13}\text{C}$ -label was introduced in  $\alpha$ -position to the

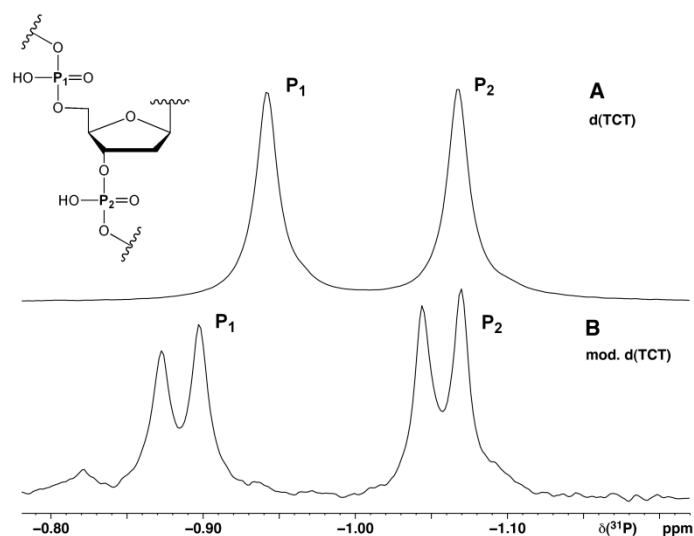
carbonyl carbon, which after carbenoid formation/insertion would directly contact the target oligonucleotide structure. The preparation of compound **40** was carried out as outlined in Scheme 41 starting from commercially available 4-(bromomethyl)bromobenzene (**38**). It was subjected to benzylic substitution with dimethylamine to give the respective 4-(dimethylaminomethyl)bromobenzene (**39**). Palladium-catalyzed C-C bond formation of that product with ethyl acetoacetate-2,4- $^{13}\text{C}_2$ , followed by an *in situ* elimination of acetate yielded intermediate **39a**. A subsequent transesterification for exchanging the alkyl ester group from ethyl to methyl (**39b**) and a nucleophilic diazo transfer on the  $\alpha$ -carbon afforded the desired compound **40** in 27 % isolated yield over four steps.



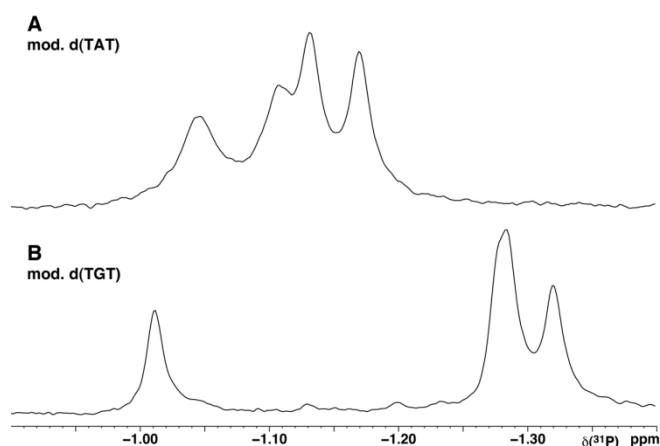
**Scheme 41** | Synthetic sequence for preparation of the  $\alpha$ - $^{13}\text{C}$ -labeled analogue of Dz1 **40**. (a)  $\text{Me}_2\text{NH}$  in EtOH (70 %). (b) i. Ethyl acetoacetate-2,4- $^{13}\text{C}_2$ ,  $\text{Pd}(\text{OAc})_2$ , 2-(di-*tert*-butylphosphino)-2'-methylbiphenyl,  $\text{K}_3\text{PO}_4 \cdot \text{H}_2\text{O}$  in toluene,  $\Delta$ ; ii. NaOMe in MeOH; iii. pABSA, DBU, MeCN (38 %).

The modification experiments were carried out on the three nested oligonucleotide trimers d(TCT), d(TAT), and d(TGT), already used in the oligonucleotide substrate scope study. Preparative separation of the two product species found for each oligonucleotide in the analytical assay was not possible, so in all cases they were further analyzed as mixtures.

$^{31}\text{P}$  NMR spectra of the total modified fractions showed two sets of phosphate signals. The resonance shifts and the peak ratios suggest they most likely correspond to the two diastereomers of the same singly-modified product (Figure 8 and 9).



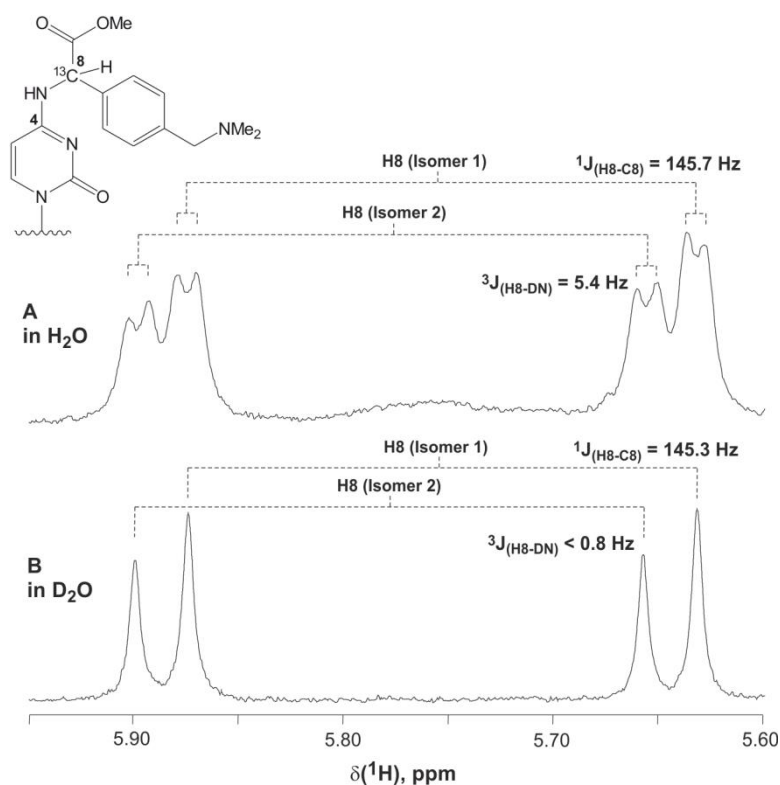
**Figure 8** |  $^{31}\text{P}$  NMR spectra of non-modified (A) and modified d(TCT) (B) with compound **40**.



**Figure 9** |  $^{31}\text{P}$  NMR spectra of d(TAT) (A) and d(TGT)(B), modified with compound **40**.

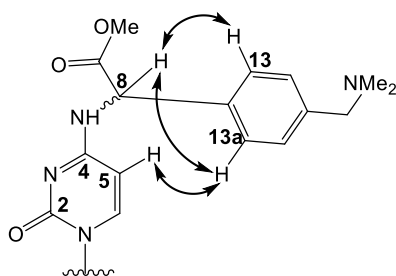
The NMR investigation of each product mixture included 1D experiments in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$ , and 2D experiments in  $\text{D}_2\text{O}$  for a complete structure assignment. A conclusive result for the position of the alkylation was first obtained for the modified d(TCT).  $^1\text{H}$  NMR analysis in  $\text{H}_2\text{O}$  of the alkylated d(TCT) revealed the following resonance pattern for its H-8: (a) two sets of signals, one for each diastereomer; (b) a large coupling to the adjacent  $\alpha\text{-}^{13}\text{C}$  ( $^1J_{\text{H8-C8}} = 145.7$  Hz); (c) a smaller coupling ( $^3J_{\text{H8-HN}} = 5.4$  Hz) to the NH of the exocyclic nitrogen of the cytosine base (N-4) (Figure 10A). When the  $^1\text{H}$  NMR was measured in  $\text{D}_2\text{O}$ , the latter coupling collapsed due to the H/D exchange of the NH ( $^3J_{\text{H8-DN}} < 0.8$  Hz). This was an unambiguous proof for the attachment of the C-8 to the exocyclic amino group of the cytosine (Figure 10B).





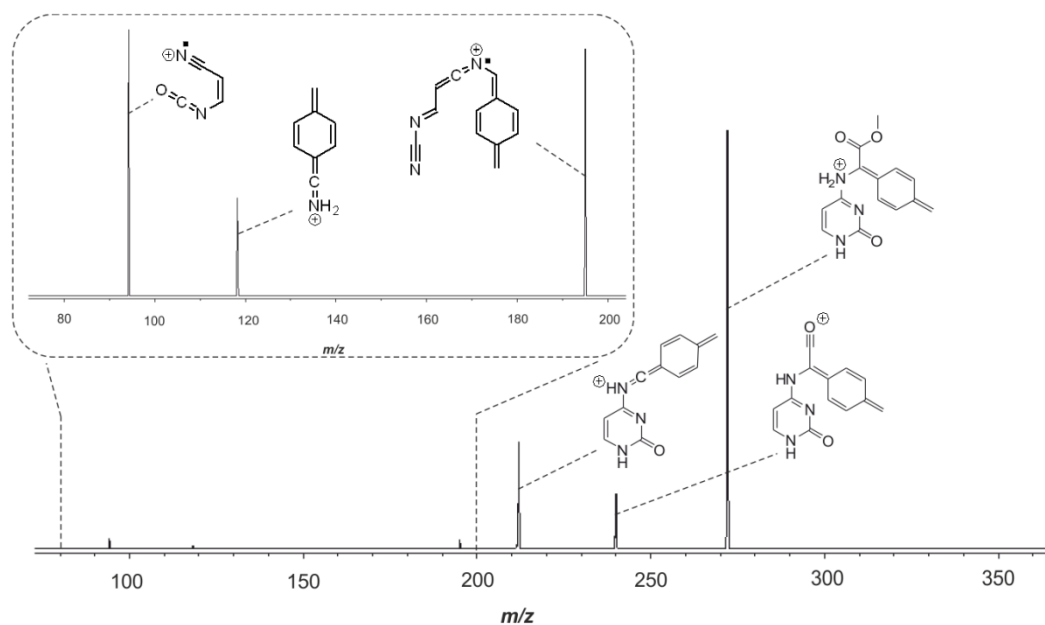
**Figure 10** |  $^1\text{H}$  NMR resonances of H-8 of modified d(TCT) in  $\text{H}_2\text{O}$  (A) and  $\text{D}_2\text{O}$  (B).

Further evidence substantiating the tag positioned on N-4 came from the ROESY spectra of the modified products. The ROE contacts between the H-5 of the cytosine base and the *ortho*-protons (H-13 and H-13a) of the aromatic portion of the tag were clearly observable (Scheme 42). In addition, no  $^3J$  correlations of the H-8 to C-2 or C-4 of the cytosine were found. Such interactions would be rather intense in case the tag was positioned at N-3.



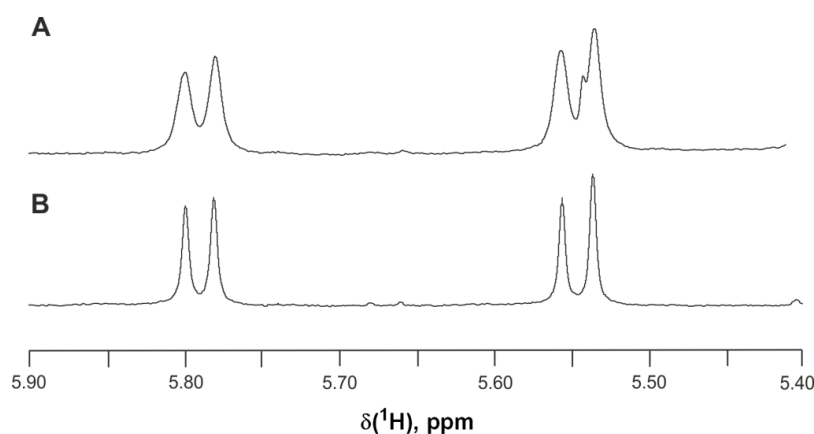
**Scheme 42** | ROE contacts of the singly-modified d(TCT) with compound **40**.

These results were in agreement with the data from the  $\text{MS}^3$  analysis of the modified d(TCT). Fragmentation of the ion of the modified base showed product ions, consistent with alkylation of the cytosine N-4 (Figure 11).



**Figure 11** | MS<sup>3</sup> fragmentation of the modified cytosine base, obtained in MS/MS fragmentation of the alkylated d(TCT).

The <sup>1</sup>H NMR data for the modified d(TAT) was rather inconclusive due to the significant broadening of the H-8 resonance both in H<sub>2</sub>O and D<sub>2</sub>O. The results for the modified d(TGT) on the other hand confirmed alkylation on the exocyclic amino group of guanine. The observed <sup>3</sup>J<sub>H8-HN</sub> coupling in H<sub>2</sub>O was smaller than that for d(TCT) and resulted in a peak broadening rather than in a clear multiplicity pattern (Figure 12). The peak width, however, decreased almost by half when the sample was measured in D<sub>2</sub>O due to the H/D exchange on the NH-group and collapse of H8-HN interactions. Establishing the alkylation site for guanine is particularly important for the present study, as it excludes the involvement of the O-6, which is also a potential target and narrows down the reaction scope to nitrogen nucleophiles only.



**Figure 12** | <sup>1</sup>H NMR resonances of H-8 of modified d(TGT) in H<sub>2</sub>O (A) and D<sub>2</sub>O (B).

### 2.3. Scope of $\alpha$ -diazocarbonyl substrates and conditions for nucleic acid alkylation

The effect of the diazo substrate structure on the modification process was examined with the  $\alpha$ -diazocarbonyl compounds shown in Table 2.

**Table 2** | Scope of diazo substrates and reaction conditions for the rhodium(II)-carbenoid alkylation with single-stranded DNA.

Entry	Diazo substrate	[Dz], mM	Buffer	Oligonucleotide conversion, % <sup>[a]</sup>
1	Dz1	50	100 mM MES, pH 6.0	56
2	Dz1	50	75 mM <i>t</i> BuNHOH, pH 6.0	52
3	Dz1	50	100 mM KPi, pH 6.0	52
4	Dz2	50	100 mM MES, pH 6.0	24
5	Dz3	50	100 mM MES, pH 6.0	48
6	Dz3	50	100 mM KPi, pH 6.5	51
7	Dz3	50	100 mM KPi, pH 7.5	50
8 <sup>[b]</sup>	Dz4	25	100 mM MES, pH 6.0	30
9 <sup>[c]</sup>	Dz5	50	100 mM MES, pH 6.0	51

Reaction conditions: 5 mM d(ATGC), 0.5 mM Rh<sub>2</sub>(OAc)<sub>4</sub>. [a] Oligonucleotide conversions were estimated from the HPLC traces of the reaction mixtures after 24 h at room temperature. [b] 50 % (v/v) ethylene glycol as a co-solvent. [c] d(TTT ATT TGT TTC TT) used as substrate instead of d(ATGC).

The alkylation reaction appears to be quite tolerant towards the diazo substrate structure, as a number of diverse functionalities were accepted, as long as the steric hindrance of the alkyl ester group was kept to a minimum. The negative effect of the increased size of the ester group was clearly demonstrated by Dz2 (**41**) in entry 4, whose reactivity towards the nucleic acid was reduced by half in comparison to Dz1 upon only a single carbon extension of the alkyl chain. On the other hand, the diazo substrates in entries 5 to 10 all showed moderate to good oligonucleotide conversions, which

were not dependent on the complexity of the aryl structure. Such reaction versatility would be particularly useful in some labeling applications, where the presence of certain chemical tags on the diazo substrate would allow further chemoselective modification of the alkylated nucleic acids. Two such examples are Dz4 (**44**) and Dz5 (**45**) (entries 8 and 9), bearing alkyne tags for decoration with azide-tagged markers by CuAAC.

Another important characteristic of the  $\alpha$ -diazocarbonyl compound is its water solubility. The standard alkylation protocol was optimized for a 10-fold excess of diazo substrate with effective concentrations of 50 mM. Tertiary or quaternary amines tend to improve the compound solubility in water, hence their use as ‘solubilizing’ motifs throughout the entire substrate series (Table 2). The addition of co-solvents in order to increase the solubility was typically avoided, as in most cases they were either incompatible with the nucleic acid causing it to precipitate (e.g. alcohols) or coordinated strongly to the dirhodium catalyst diminishing its activity (DMSO). In the case of Dz4 ethylene glycol was used as a co-solvent in a 1:1 mixture with the aqueous phase (entry 8). It should be noted, however, that in a separate experiment the addition of the same amount of ethylene glycol decreased the yield of alkylated d(A)<sub>4</sub> by half (from 39 to 20 % conversion of the oligonucleotide). In that sense the solubility of the alkyne-tagged diazo substrate Dz4 was improved by altering the compound’s structure. The introduction of a second piperazine ring and a short PEG-chain afforded a compound fully soluble at a concentration of 50 mM allowing for satisfying yields of alkylation (Entry 9).

The alkylation reaction was typically carried out at pH 6.0, as this pH insures the stability of the nucleic acid species and in the same time is optimal for the diazo substrate solubility. Experiments with the *N*-permethylated Dz3 (**42**) showed that the alkylation efficiency remained practically the same in the pH range up to 7.5 (Table 2, entries 5-7). The nature of the buffering substance appeared to have very little effect on the reaction at pH 6.0 (Table 2, entries 1-3). The addition of *N*-*tert*-butylhydroxylamine, which was previously discussed as beneficial in the case of protein alkylation,<sup>55</sup> seemed to have no effect on the reaction with nucleic acids (Table 2, entry 2).

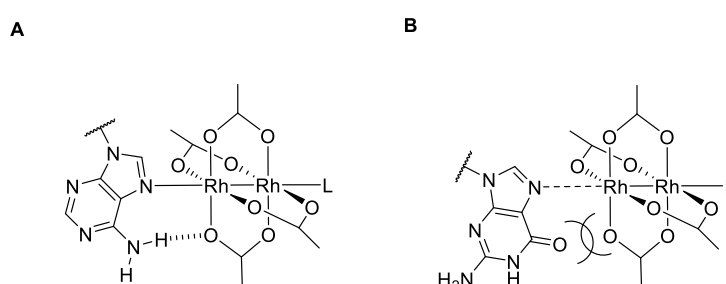
#### **2.4. Mechanistic rational for the Rh(II)-carbenoid nucleic acid modification**

A mechanistic assessment of the rhodium(II) alkylation system could prove difficult due to the complexity of the reaction substrates, the possibility for multiple coordination states of the rhodium(II) catalyst and the presence of a side-reaction of O-H insertion, taking place concurrently and accounting for over 90 % of the diazo substrate conversion.

Such mechanistic assessment was attempted for the alkylation of the three nested DNA trimers d(TCT), d(TAT) and d(TGT) with Dz1. It should be noted that both the oligonucleotides and the diazo substrate contain a number of Lewis basic groups, which could potentially coordinate to the metal center altering its catalytic properties. Attempts to carry out an O-H insertion reaction with Dz1 and 1

mol% rhodium(II) tetraacetate in aqueous buffer in the absence of an oligonucleotide delivered less than 35 % overall conversion of the diazo substrate over 20 h at room temperature, while in the presence of d(TCT) or d(TAT) complete conversion of Dz1 to the respective N-H and O-H insertion products was observed. The protective role of the oligonucleotide substrate clearly indicate the presence of ligand effects, which are responsible for the drastic changes in the rhodium(II) catalytic activity.

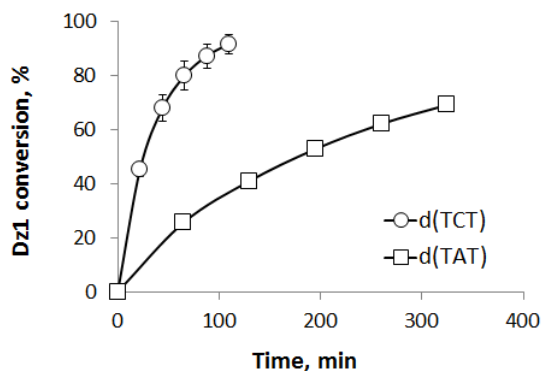
The coordination of nucleoside derivatives, as well as oligonucleotides and DNA and RNA to rhodium(II) is also a well-known phenomenon. The coordination of adenine and adenosine in particular, first described by Farrell for rhodium(II) tetraacetate,<sup>93</sup> takes place at the axial position of the bimetal core through N-7 of the nucleobase (Scheme 43A).



**Scheme 43** | Axial coordination of an adenine (A) or a guanine derivative (B) to  $\text{Rh}_2(\text{OAc})_4$ .

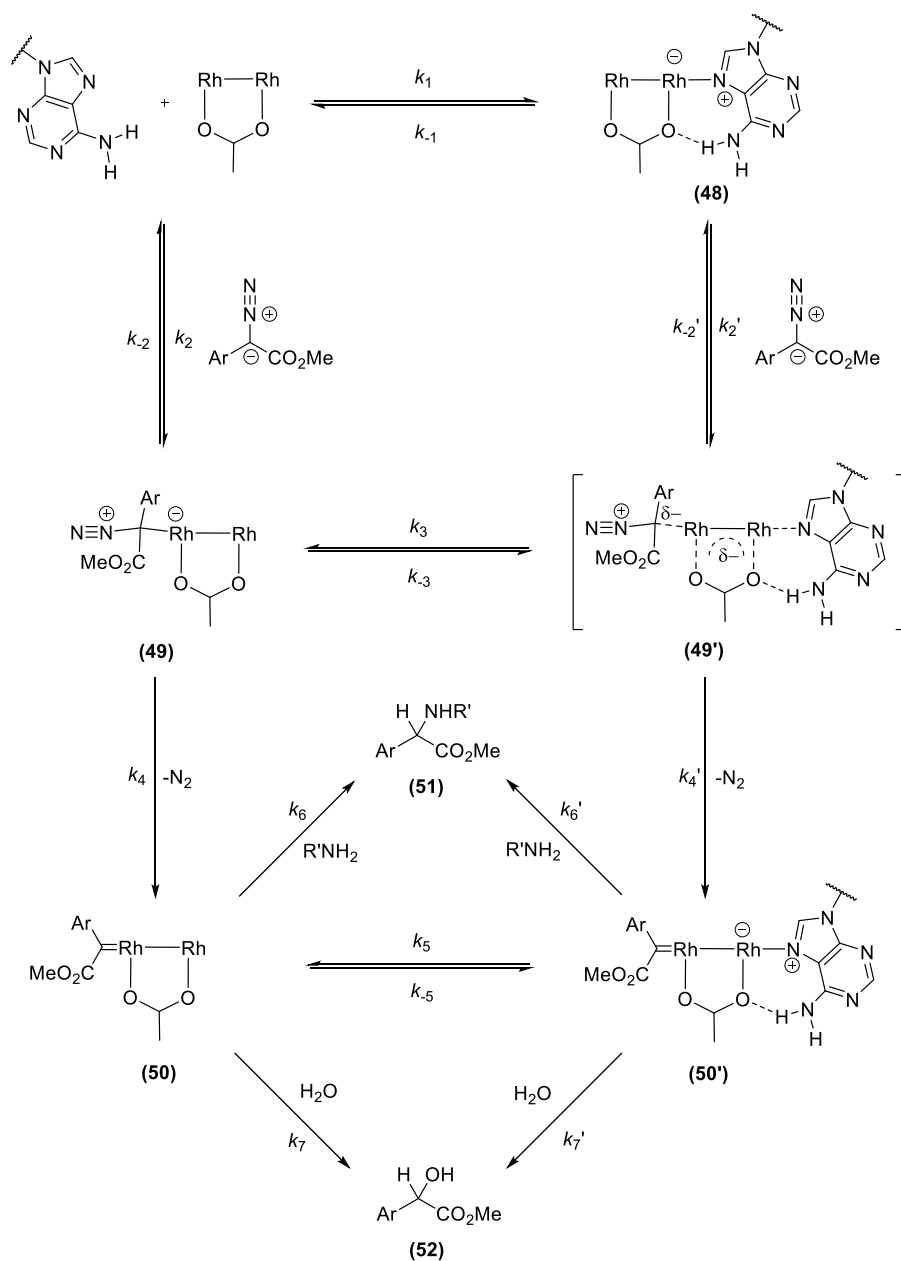
The interaction is additionally stabilized by hydrogen bonding between the exocyclic amino group of the adenine (N-6) and the carboxylate oxygen of the acetate ligand. A similar interaction is to be expected for the cytosine base, although the literature data about its axial coordination in rhodium(II) tetracarboxylates is absent.

The complex formation between d(TCT) and d(TAT) and the dirhodium species was confirmed by kinetic studies. Experiments with d(TCT) and d(TAT), rhodium(II) tetraacetate and Dz1 showed significantly different rates of diazo substrate disappearance, as  $\tau_{1/2}$  for d(TCT) was around eight times smaller than that for d(TAT) (Figure 11).



**Figure 11** | Time curves for the overall conversion of Dz1 in the presence of  $\text{Rh}_2(\text{OAc})_4$  and d(TCT) or d(TAT).

Having in mind that the oligonucleotide N-H insertion reaction accounts for less than 3 % of the total Dz1 conversion, the nature of the nucleobase should not significantly affect the rate of the O-H insertion process, unless the oligonucleotide itself would be able to bind to the catalyst and alter its activity. A possible mechanistic model involving oligonucleotide pre-coordination to the rhodium tetraacetate catalyst is depicted in Scheme 44 for the case of d(TAT). The process sequence that would take place also in the absence of a coordinating nucleobase would start with coordination of the diazo substrate to the rhodium catalyst, yielding complex **49**, followed by loss of nitrogen to give carbene **50**, which then reacts with an amine or water to afford the final insertion products **51** and **52** respectively. An additional axial coordination of the adenine nucleobase could in principle take place at any stage of the process, hence intermediates **48**, **49'** and **50** should be considered. According to the findings of Pirrung and Morehead if one axial site on the dirhodium core is already occupied the second axial site should have a significantly decreased binding capability, thus diminished catalytic activity.<sup>61,62</sup> In that sense the formation of species **49'** from **48** should be rather unfavorable but still possible if the coordination of the  $\alpha$ -carbon of the diazo substrate were accompanied by a partial ligand dissociation at the distal axial site, as well as in the equatorial positions. On the other hand the increased electron density of the bimetal core in **49'** should favor the back-donation from the dirhodium to the coordinated diazo compound, thus facilitating the extrusion of nitrogen and the formation of carbenoid **50'**. Finally, nucleobase coordination or dissociation after the carbene has formed should be considered only if these were significantly faster than the N-H and O-H insertion reactions consuming that very carbene.

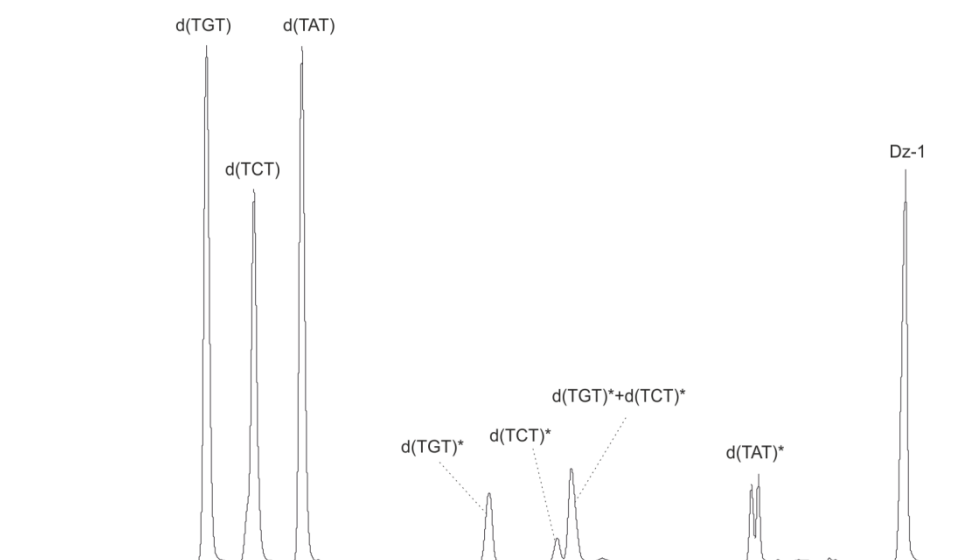


**Scheme 44** | Putative mechanistic model of D21 conversion in the presence of  $\text{Rh}_2(\text{OAc})_4$  and d(TAT).

The alkylation of d(TGT) presented quite an interesting case. Attempts to alkylate d(TGT) alone were unsuccessful. Addition of an equimolar amount of d(TAT) to the reaction mixture alleviated that problem, yielding efficient alkylation of both trimers. This is another evidence of the protective role of the adenine axial coordination. The inability of guanine to form stable complexes to the axial position of rhodium(II) could be explained with repulsive interaction, occurring between the O-6 of guanine and the oxygen of the carboxylate equatorial ligand (Scheme 43B).

The individual reactivity of the nucleobases in the N-H insertion step was determined in a competition experiment with the three trimers d(TCT), d(TAT), and d(TGT) (Figure 12). The propensity for

alkylation decreased in the order: G>A>C with oligonucleotide conversion 25, 18, and 12 % respectively. It should be noted, that nucleobase targeting appears to be sequence-specific, as for example the fully reactive base cytosine remained completely untouched in the d(ATGC) oligonucleotide.



**Figure 12** | HPLC trace of the mixed alkylation of d(TCT), d(TAT), and d(TGT) with Dz1 in the presence of Rh<sub>2</sub>(OAc)<sub>4</sub>. The total peak area of modified d(TCT) is determined from the peak area of the resolved diastereomer assuming equal amount for the other diastereomer.

## 2.5. Structure selectivity of the Rh(II)-catalyzed alkylation: experiments with double-stranded DNAs and RNAs

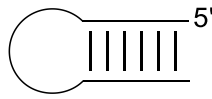
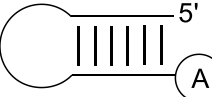
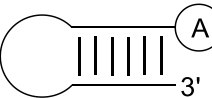
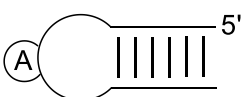
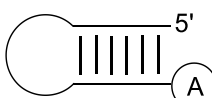
Double-stranded DNAs and RNAs are an inseparable part of the nucleic acid structure variety. In fact only the actively replicated and transcribed DNA could be found in the single-stranded form. The situation with RNA is even more complex, as all types tend to form diverse secondary structures with alternating single- and double-stranded regions. In that sense the ability of the rhodium(II) alkylation system to target nucleic acid duplex motifs is an important characteristic concerning the method scope and potential practical application.

The double-strand alkylation experiments were performed on model DNA and RNA hair-pins (Table 3). The use of hair-pins has several advantages over nucleic acid duplexes pre-formed by mixing equimolar amounts of the respective complementary strands. It requires the synthesis and purification of only a single strand and, when properly designed, it insures the quantitative formation of the desired duplex species only. Several conceptually different structures were generated by only small changes in



the base sequences. All of the hair-pins used contained a small loop of five nucleobases and a short duplex portion (6-10 bp).

**Table 3** | Scope of the rhodium(II)-carbenoid alkylation with double-stranded DNA and RNA oligonucleotides.

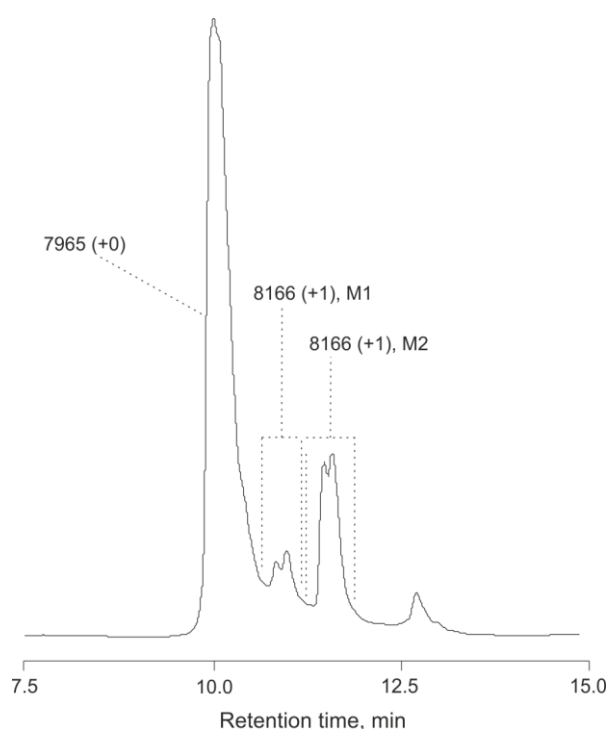
Entry	Hair-pin substrate	Structure	Conversion, % <sup>[a]</sup>
1	HP1 d(CGAACGTTTTTCGTTTCG)		0
2	HP2 d(CGAACGTTTTTCGTTTCGA)		21
3 <sup>[b]</sup>	HP3 d(ACGGAATTCCGTTTTTCGGAATTCCG)		19
4	HP4 d(CGAACGTTATTCGTTTCG)		6 22 <sup>[c]</sup>
5	HP5 r(CUAGCAUUUUUUCGUAGA)		19 33 <sup>[c]</sup>

Reaction conditions: 5 mM oligonucleotide, 50 mM Dz1, 0.5 mM Rh<sub>2</sub>(OAc)<sub>4</sub>, 100 mM MES buffer, pH 6, 24 h at room temperature. [a] Oligonucleotide conversion determined from the HPLC analysis of the reaction mixtures. [b] This sequence contains a restriction site for *Eco*RI. Alkylation at 5'-adenine overhang confirmed by MS/MS analysis of the restriction digest with *Eco*RI. [c] Conversion after addition of a second portion of Dz1 and 24-h incubation.

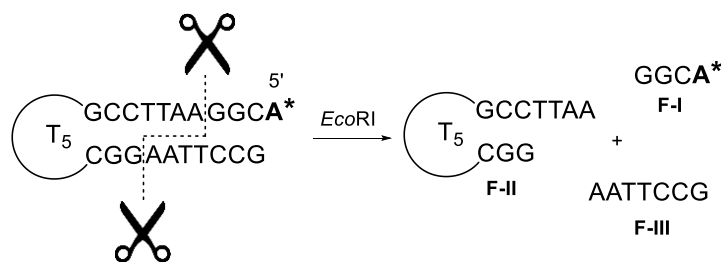
Alkylation of the hair-pins was attempted with Dz1 under the conditions already worked out for the single-stranded nucleic acids. Hair-pin structure 1 (HP1, Table 3, Entry 1), containing only unreactive thymine bases in the loop and no overhangs in the stem region was completely inert to alkylation with rhodium carbenoids, showing that bases, engaged in double strands, were *not* targeted. Once a reactive base (adenine) was introduced outside the duplex region as a 3'- (HP2, Entry 2) or a 5'-overhang (HP3, Entry 3), or as a base in the loop region (HP4, Entry 4) modest conversions of the oligonucleotide to the monoalkylated products were observed. The reactivity of HP4 with adenine in the loop region was significantly lower than that for HP2 and HP3, and required the addition of a second portion of Dz1 and longer incubation times (48 h) to achieve the same oligonucleotide conversion. This was most likely due unfavorable sterics of the adenine in the loop structure. The propensity for alkylation of the RNA hair-pin HP5 with a 3'-adenine overhang was comparable to that of the DNA counterparts (Entry 5).

The position of the alkylation, occurring outside of the duplex region, was confirmed with HP3, which contained a restriction site for the restrictase *EcoRI* inside the stem. Upon treatment with Dz1 in the presence of Rh(II) the formation of two sets of alkylation products M1 and M2 was observed (Figure 13). The alkylation products were isolated by micro-preparative RP-HPLC and digested with *EcoRI* to release fragments F-I to III shown on Scheme 45. HPLC-MS confirmed the presence of the singly-modified fragment F-I and the unmodified fragment F-II in the restriction digest. Further MS/MS analysis of the monoalkylated F-I revealed the modification had taken place on the adenine base of the 5'-overhang (Figure 14).

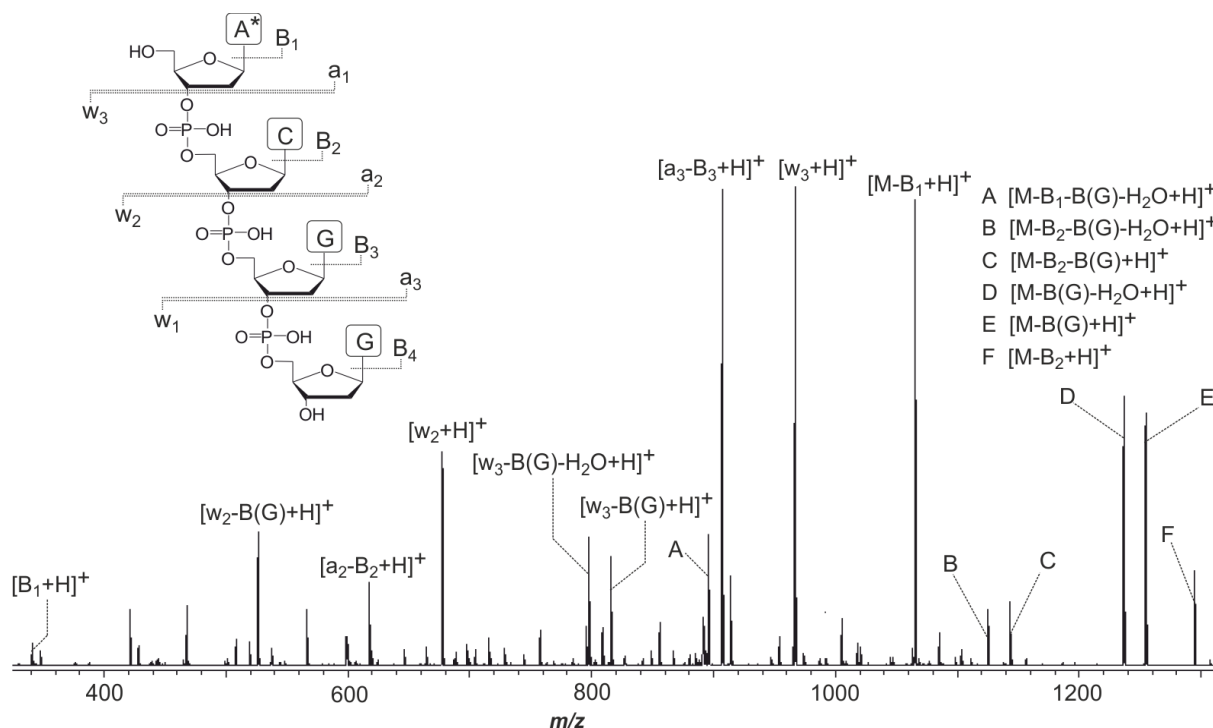
The ability of the rhodium carbenoids to efficiently select for single- over double-stranded nucleic acid motifs is an important feature of the method and could be of direct practical significance. Such structure selection could be utilized for site-specific alkylation of single strands using duplex formation as a protection strategy. The target sequence is mixed with one or several shorter complementary chains, which after duplex formation would leave only the desired bases exposed, allowing site-specific delivery of the desired modification. Subsequent purification under denaturing conditions should yield the target species with the desired modification in place.



**Figure 13** | HPLC trace of the Rh(II)-catalyzed alkylation of HP3 with Dz1. The product masses and the number of alkylations per oligonucleotide molecule (in parenthesis) are indicated for each peak.



**Scheme 45** | Restriction fragments of HP3 upon treatment with *EcoRI*. The alkylated 5'-end adenine is indicated with an asterisk.



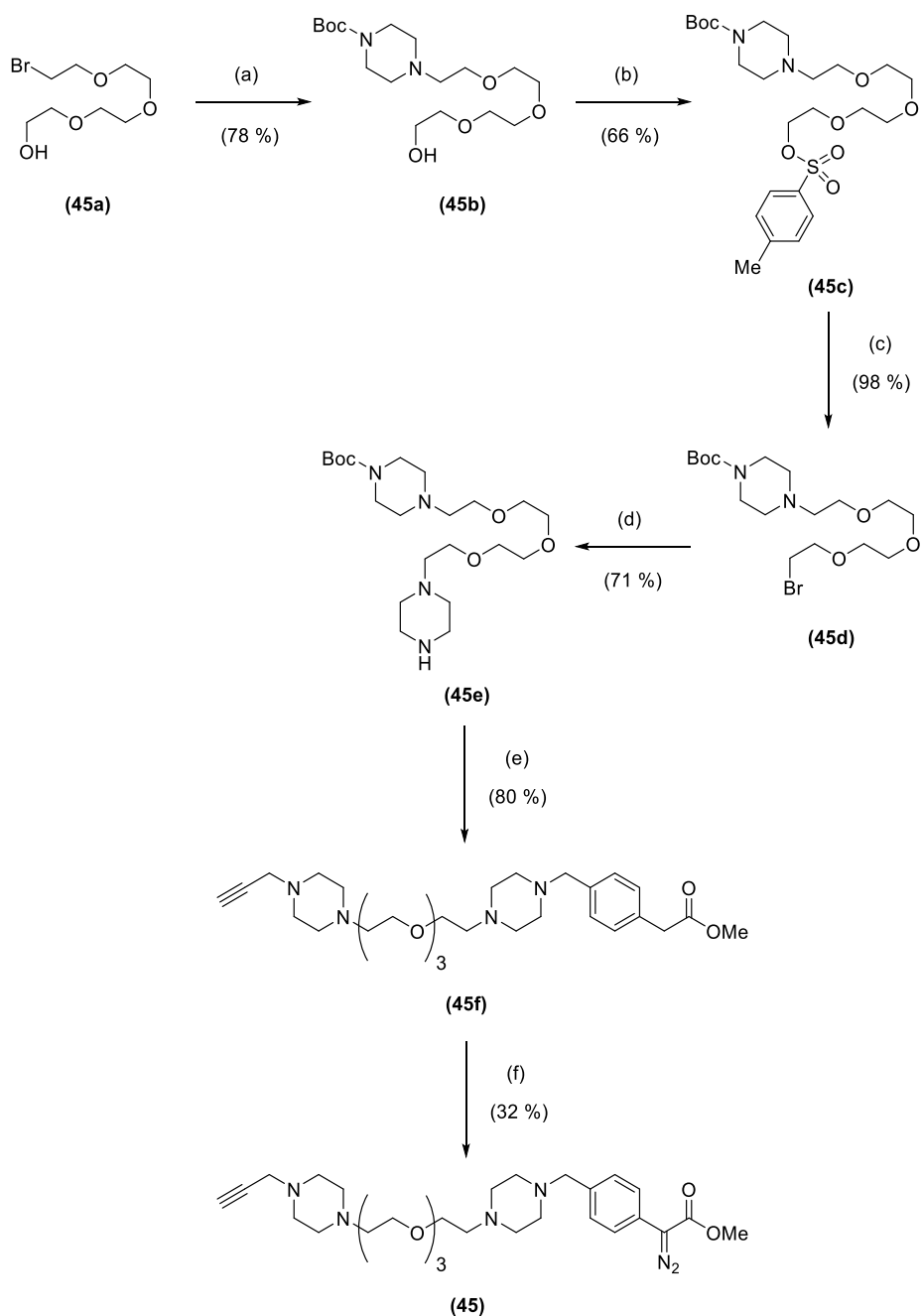
**Figure 14** | Tandem MS fragmentation of the 5'-end fragment d(A\*CGG) from the *EcoRI* restriction of the modified HP3. The alkylated 5'-end adenine is indicated with an asterisk.

## 2.6. Nucleic acid labeling applications

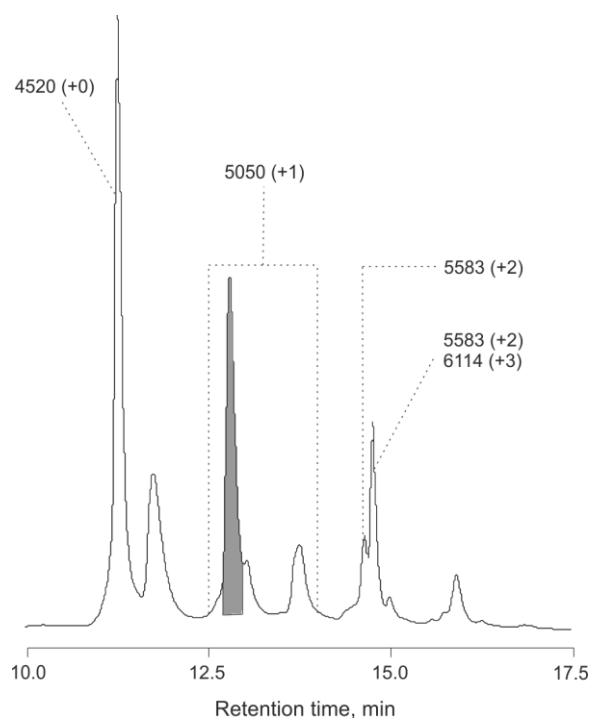
The direct introduction of label moieties already in the diazo substrate structure proved to be tedious and resulted in sterically crowded compounds which were poorly soluble in water, displaying less than optimal alkylation properties as in the example of Dz7. In a more facile divergent strategy only a single chemical tag (alkyne group) was installed on the diazo substrate used for nucleic acid alkylation. Diversification of the product structure was done on a later stage by chemoselective functionalization of the alkyne tag by CuAAC with appropriately labeled azides.

All labeling experiments were carried out with Dz5 discussed earlier, which was prepared as outlined in Scheme 46. The dipiperazine tri(ethyleneglycol) spacer arm was constructed starting from 1-

bromotetra(ethyleneglycol) (**45a**), then installed on the diazo core structure by benzylic substitution in compound **36**. Boc-group removal under neutral conditions, followed by N-propargylation of the terminal piperazine afforded compound **45f**, which after nucleophilic diazo transfer on the  $\alpha$ -carbonyl carbon yielded Dz5 (**45**).

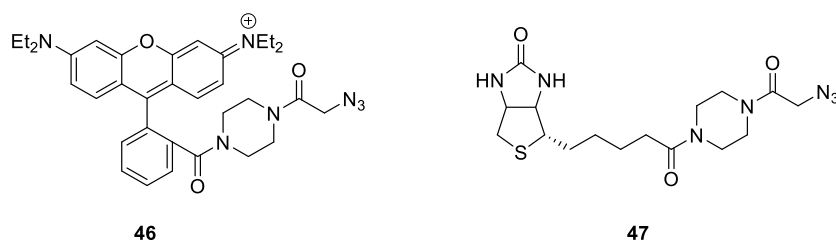


**Scheme 46** | Synthetic sequence for preparation of Dz5 (**45**): (a) Boc-piperazine, K<sub>2</sub>CO<sub>3</sub>. (b) pTsCl, Et<sub>3</sub>N, DMAP. (c) LiBr. (d) Piperazine, K<sub>2</sub>CO<sub>3</sub>. (e) i. **36**, K<sub>2</sub>CO<sub>3</sub>; ii. TMSOTf, 2,6-lutidine; iii. Propargyl bromide, K<sub>2</sub>CO<sub>3</sub>. (f) pABSA, DBU.



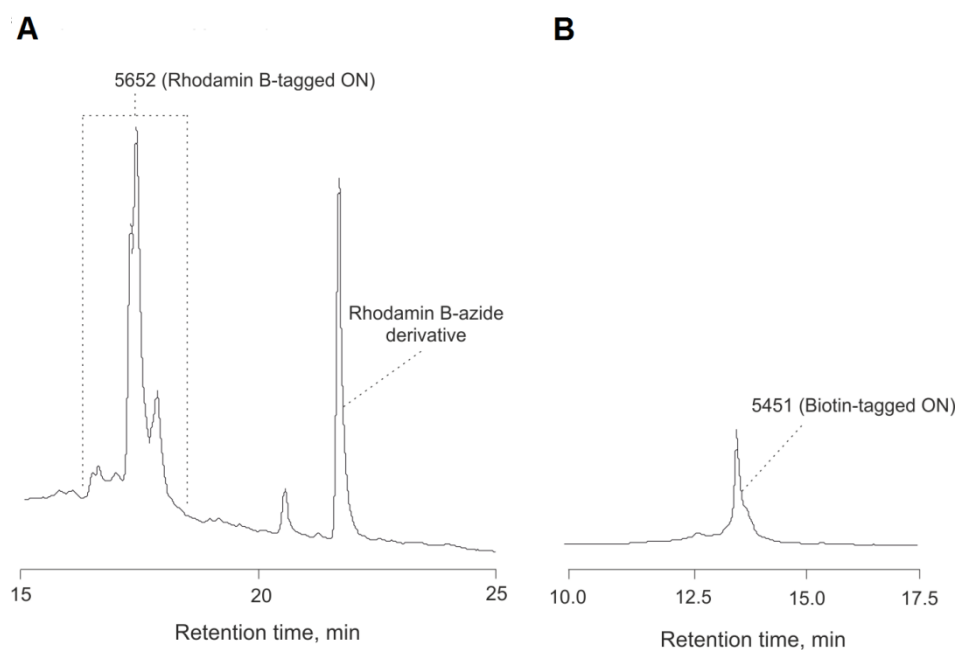
**Figure 15** | HPLC-separation of the modification reaction mixture of d(TTT ATT TGT TTC TTT) with propargylated Dz5. The isolated fraction of main monoalkylation product is indicated by the solid peak area. The product masses determined by MALDI-TOF and the number of alkylations per oligonucleotide molecule (in parenthesis) are indicated for each peak.

Modification with Dz5 was first done with the model DNA 15-mer d(TTT ATT TGT TTC TTT). A number of singly-, doubly- and even triply-modified products were formed at a 51 % conversion of the oligonucleotide as judged by the HPLC-analysis of the reaction mixtures (Figure 15). The main monoalkylated product was isolated preparatively and further modified with rhodamine B and biotin-tagged azides **46** and **47** respectively under the conditions described by Finn et al.<sup>86</sup> to afford the respective ‘click’-products (Figure 16).



The utility of the nucleic acid alkylation with rhodium(II) carbenoids was then demonstrated by the fluorescent tagging of T7 promoter primer for preparation of labeled PCR amplicons (Scheme 47). T7 promoter primer with a 5'-d(C)<sub>4</sub>-extension (d(C)<sub>4</sub>-T7PP) was first alkylated with Dz5 in the presence of Rh<sub>2</sub>(OAc)<sub>4</sub>, and then coupled to rhodamine B azide *via* CuAAC. The identity of the products was confirmed by MALDI-TOF and their viability as functioning primers – in a PCR assay with a 275-bp

test template sequence, followed by an agarose gel electrophoretic analysis of the PCR mixtures. Successful labeling was first indicated by the presence of a fluorescent PCR amplicon already before DNA visualization with ethidium bromide (Lane 5a).



**Figure 16** | HPLC-separation of the ‘click’ reaction of the main monopropargylated d(TTT ATT TGT TTC TTT) with rhodamine B (A) and biotin azide (B).

The ethidium bromide soak showed amplification comparable to that with the control primer (Scheme 47, Lane 2b and 5b). The propargylated d(C)<sub>4</sub>-T7PP, both purified (Lane 3b) or directly from the alkylation reaction mixture (Lane 4b) also delivered the desired amplification product. Despite the random manner of base modification, the described approach was able to produce fluorescently labeled PCR product of the correct length and in satisfying yields. It is likely that only modifications close to the primer 5'-end would result in efficient template binding and amplification of the desired sequence. Since the PCR primers are normally added in a large excess, even the small fraction of ‘properly’ modified at the 5'-end primer molecules should be sufficient for a robust amplification of the target sequence. The ease of label incorporation and the possibility to tag virtually *any* practically important primer sequence makes the method a viable alternative to the standard pre-synthetic nucleic acid modification approaches.



## 2.7. Conclusions and future directions

A novel alkylation method for short and medium-sized DNA and RNA oligonucleotides using rhodium(II)-stabilized carbenoids has been developed. The method allows for an efficient and clean conversion of the nucleic acid substrate to a number of singly-, doubly- and triply-modified species. It utilizes a simple rhodium(II) precursor,  $\text{Rh}_2(\text{OAc})_4$ , and a stabilized  $\alpha$ -diazocarbonyl compound for the *in situ* generation of the active rhodium(II) carbenoids under mild conditions - aqueous medium at ambient temperature and pH close to the physiologically normal one.

Most importantly, the method provides access exclusively to modification of the exocyclic amino groups in adenine, guanine and cytosine without affecting the more nucleophilic heterocyclic nitrogen atoms or O-6 in guanine. The regiospecific modification of the exocyclic amines is achieved in a structure-selective manner in single-stranded nucleic acid species only, allowing the Watson-Crick base pairing to be used as a protective motif. Several examples of nucleic acid labeling with fluorescent and biotin tags demonstrate the utility of the method.

The future work on the rhodium(II) carbenoid system could be continued in several parallel directions:

- (a) Expanding the nucleic acid scope – targeting of longer nucleic acid species (100 bp to several kb). This would require development of analytical tools of higher precision and sensitivity in order to detect even small changes in the context of a longer DNA or RNA molecule. In this sense high resolution mass spectrometry or labeling with specific tags (e.g. biotin) would prove invaluable in creating more tractable test systems;
- (b) Expanding the scope of  $\alpha$ -diazocarbonyl substrates:
  - a. Diazo substrates carrying certain tag moieties;
  - b. More activated compounds in order to increase the N-H/O-H insertion ratio, thus improve the yield of alkylation (e.g. ethyl diazoacetate);
- (c) An in-depth mechanistic investigation of the system. The development of a correct mechanistic model for a system with both substrates functioning as potential ligands proved to be quite challenging. A reaction assembly with a re-designed diazo substrate that can coordinate only through the electron-rich  $\alpha$ -carbon atom would simplify a further mechanistic treatment. An investigation of the pre-catalytic coordination states of the dirhodium catalyst would be crucial for such a mechanistic description, as they appear to be detrimental for the catalytic process:
  - pre-catalytic complex between rhodium(II) and the nucleobase of the substrate – type of coordination (axial or equatorial), stability of the complex, possibilities for axial ligand exchange;
  - possibility for equatorial ligand exchange and guanine coordination on the reaction time scale and conditions;



- a theoretical/computational study – calculation of the individual free energies for a small collection of possible pre-catalytic complexes and reaction intermediates. This would be particularly valuable in the assessment of the potential rhodium(II) coordination states in solution;
- kinetic investigation of the reaction – determination of the rate-limiting step – loss of nitrogen/carbenoid formation or nucleophilic attack on the carbenoid/proton shift in the insertion step;
- influence of the potential coordination states of the catalyst on the sequence-specificity of the reaction.

### Chapter 3. Copper(I)-catalyzed insertion reactions in aqueous media

The main goal of the work presented in this chapter is to develop a methodology for the alkylation of nucleic acids and simple anilines with copper(I)-stabilized carbenes in aqueous media. The use of this approach alone or in a one-pot set-up with CuAAC would provide access to variety of potentially significant product structures.

#### 3.1. Alkylation of nucleic acids. Optimization of the reaction conditions and scope with single- and double-stranded DNAs and RNAs

The successful implementation of the rhodium(II) carbenoid system for nucleic acid alkylation hinted towards the possibility other transition metals to be used for the same purpose. A small screen of several transition metal species in an alkylation reaction with the model oligonucleotide d(ATGC) and Dz1 identified copper(I) as the most promising candidate for such a catalytic application (Table 4).

**Table 4** | Metal catalyst screening for oligonucleotide alkylation.

Entry	Metal	Metal source	Number of alkylation products	Oligonucleotide conversion (%) <sup>[a]</sup>
1	Cu(II)	CuSO <sub>4</sub>	8	11
2	Ag(I)	AgBF <sub>4</sub>	6	3
3	Co(II)	CoCl <sub>2</sub>	5	3
4	Ni(II)	NiCl <sub>2</sub>	5	3
5	Fe(III)	FeCl <sub>3</sub>	5	4

Reaction conditions: 5 mM oligonucleotide, 1 mM metal source, 50 mM Dz1, 100 mM MES, pH 6.0, 24 h, room temperature. [a] Determined by HPLC-analysis of the reaction mixture.

Initial tests with several copper sources and ligands of both +1 and +2 oxidation states of the metal outlined a few important features of the reaction system (Table 5). The addition of a reducing agent (sodium ascorbate) to preserve the copper(I) oxidation state proved to be quite beneficial for the alkylation reaction. This was in agreement the earlier findings of Kochi<sup>78</sup> and Teyssie,<sup>60</sup> who had shown that copper(I) was the true catalytically active species in copper carbenoid reactions. In the cases, where copper(II) was used, it had been reduced *in situ* by the  $\alpha$ -diazocarbonyl compound present to afford the catalytically active copper(I) species. The two strongly binding bidentate ligands 1,10-phenanthroline and 2,2'-bipyridyl seemed to diminish the alkylation efficacy (Table 5, Entries 4,

5, 8, 9). This was most likely due to a tighter binding to the copper(I) center and decreasing its affinity for the diazo substrate. The triazole ligand THPTA, on the other hand, increased the yield of alkylated product. It likely serves the dual roles of stabilizing the copper(I) species in aqueous media and scavenging the reactive oxygen species generated by the metal in the presence of ascorbate.<sup>86</sup> Such products, which would otherwise cause oxidative damage of the nucleic acid species are likely directly intercepted by sacrificial ligand molecules on their way out of the inner copper coordination sphere.

**Table 5** | Optimization of the reaction conditions for the copper-catalyzed alkylation of d(ATGC) with Dz1

Entry	Copper source	Ligand	Addition of ascorbate <sup>[a]</sup>	Number of alkylation products		Oligonucleotide conversion, % <sup>[b]</sup>
				(+1)	(+2)	
1	CuSO <sub>4</sub>	-	-	>7	-	14
2	Cu(OTf) <sub>2</sub>	-	-	7	-	13
3	CuSO <sub>4</sub>	THPTA	-	>7	-	18
4	CuSO <sub>4</sub>	1,10-Phenanthroline	-	7	-	6
5	CuSO <sub>4</sub>	2,2'-Bipyridyl	-	7	-	6
6	CuSO <sub>4</sub>	-	+	7	>4	47
7	CuSO <sub>4</sub>	THPTA	+	7	>4	52
8	CuSO <sub>4</sub>	1,10-Phenanthroline	+	7	>4	36
9	CuSO <sub>4</sub>	2,2'-Bipyridyl	+	7	>3	30

Reaction conditions: 5 mM d(ATGC), 1 mM copper(II) source, 1 mM ligand, 50 mM Dz1, 100 mM MES buffer, pH 6.0, 24 h, room temperature. [a] Copper(I) was generated *in situ* by addition of 50 mM sodium ascorbate. [b] Determined by HPLC-analysis of the reaction mixture.

In its final variant the alkylation conditions converged with those for the CuAAC, which would later allow carrying out both reactions concurrently. The potential of the copper(I) carbenoids for nucleic acid alkylation was demonstrated on a small series of short and medium-sized oligonucleotides (Table 6). Alkylation of the model oligonucleotide d(ATGC) was a lot more efficient than with the rhodium(II) system, as conversions of up to 70 %, accompanied by the formation of over twelve alkylation products were observed (Table 6, Entry 1). A more detailed HPLC/MS-analysis of the reaction mixtures revealed the presence of even triply-modified products (Figure 17).

The three trinucleotides d(TCT), d(TAT), and d(TGT), were used to probe the reaction's regioselectivity and the behavior of the individual nucleic acid structural elements. Apart from the almost two-fold increase in the oligonucleotide conversion, the alkylation pattern of d(TAT) with copper(I) was identical to that with rhodium(II). The modification of d(TGT) on the other hand showed some considerable differences to the rhodium(II)-catalyzed reaction.

**Table 6** | Scope of copper(I) nucleic acid alkylation with Dz1.

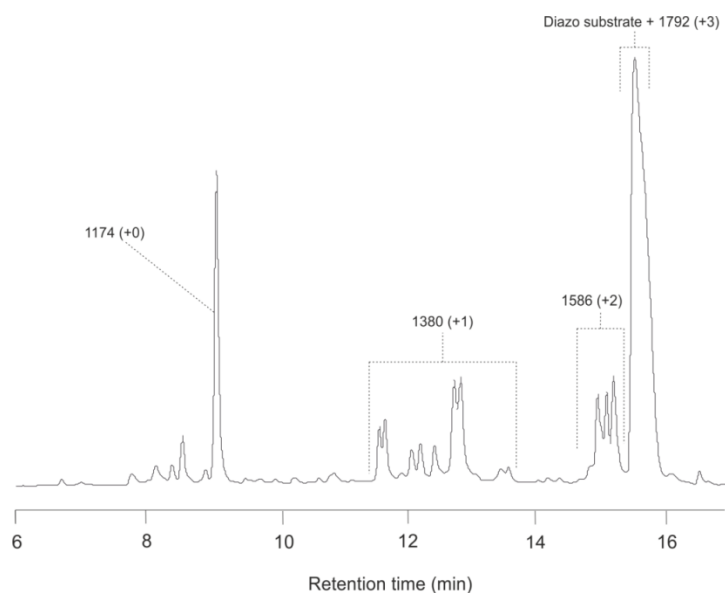
Entry	Oligonucleotide sequence	Number of alkylated species			Oligonucleotide conversion, % <sup>[a]</sup>	Retained oligonucleotide, % <sup>[b]</sup>
		(+1)	(+2)	(+3)		
1	d(ATG C)	7	4	1	70	>77
2	d(TAT)	2	-	-	37	>87
3	d(TGT)	4	1	-	46	>74
4	d(TCT)	4	-	-	27	>85
5	d(TTT T)	7	1	-	19	>76
6 <sup>[c]</sup>	d(ATG)	3	2	2	61	-
7	r(ACU GCU C)	>11	-	-	67	-
8 <sup>[d][e]</sup>	d(CGA ACG TTT TTC GTT CG)	2	-	-	9	-
9 <sup>[d][e]</sup>	d(CGA ACG TTT TTC GTT CGA)	4	-	-	16	-
10 <sup>[d][e]</sup>	d(CGA ACG TTA TTC GTT CG)	6	-	-	16	-

Reaction conditions: 5 mM oligonucleotide, 0.5 mM CuSO<sub>4</sub>, 2.5mM THPTA, 50 mM Dz1, 10 mM sodium ascorbate, 100 mM MES buffer, pH 6, 24 h, room temperature. [a] Determined by HPLC-analysis of the reaction mixture. [b] Determined for entries 1-5: total peak area of all oligonucleotide species after the alkylation reaction as a percentage of the peak area of the starting oligonucleotide. [c] 5 mM sodium ascorbate. [d] Reaction time 48 h. [e] Forms a hair-pin structure in solution.

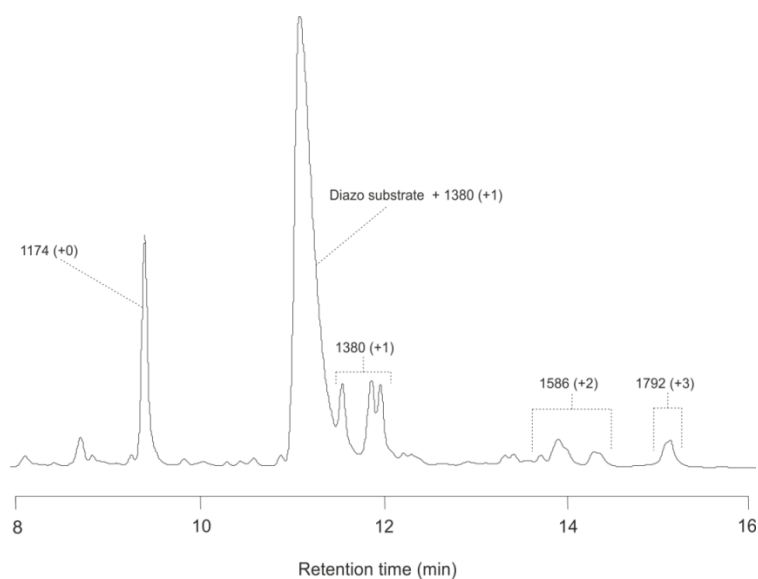
First, d(TGT) was efficiently targeted in the absence of other oligonucleotides. Second, in addition to the first set of monoalkylated species, already seen with rhodium(II), the formation of a second set of reaction products, as well as some doubly-alkylated oligonucleotide was observed (Figure 18A). A second pair of alkylated oligonucleotides was also found in the d(TCT) reaction, although their amount was significantly lower than in the case of d(TGT) (Figure 18B). Tandem MS analysis of these products revealed the second reaction site was again located on the nucleobase. The ion for the doubly-alkylated guanine base was directly observed in the MS/MS analysis of the product from the d(TGT) doubly-alkylated fraction (Figure 19). The second modifications most likely occurred at the heterocyclic nitrogen atoms N-3 in cytosine and N-1 in guanine, or on the carbonyl oxygens, O-2 and O-6 respectively, in the enol forms of the bases (Scheme 48). Attempts for isolation of these products from the d(TCT) reaction mixture were unsuccessful, as they were not sufficiently stable and seemed to convert to the other two species, already seen in the rhodium(II) alkylation.

The next small oligonucleotide tested was d(ATG). Alkylation with Dz1 in the presence of copper(I) resulted in several singly-, doubly-, and triply-modified products, consistent with monoalkylation of the adenine and mono- and double alkylation of the guanine (Table 6, Entry 6).

**A**

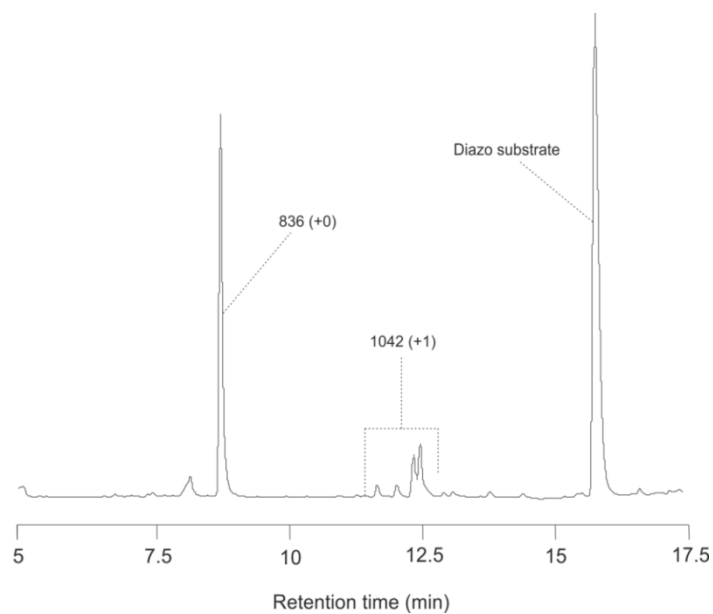


**B**

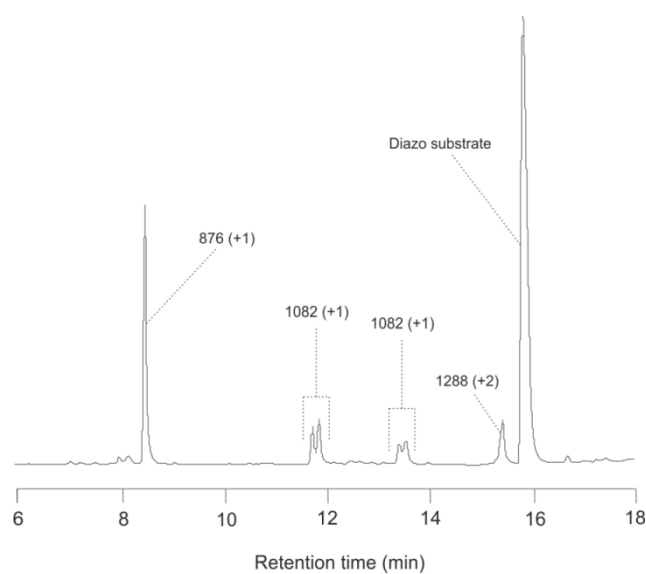


**Figure 17** | HPLC-separation of the reaction mixture from the copper(I)-catalyzed alkylation of d(ATGC) with Dz1. HPLC-separation at pH 7.0 (A) and 6.5 (B). The number for each peak indicates the mass from ESI-MS the number in parenthesis indicates the number of alkylations per molecule of oligonucleotide.

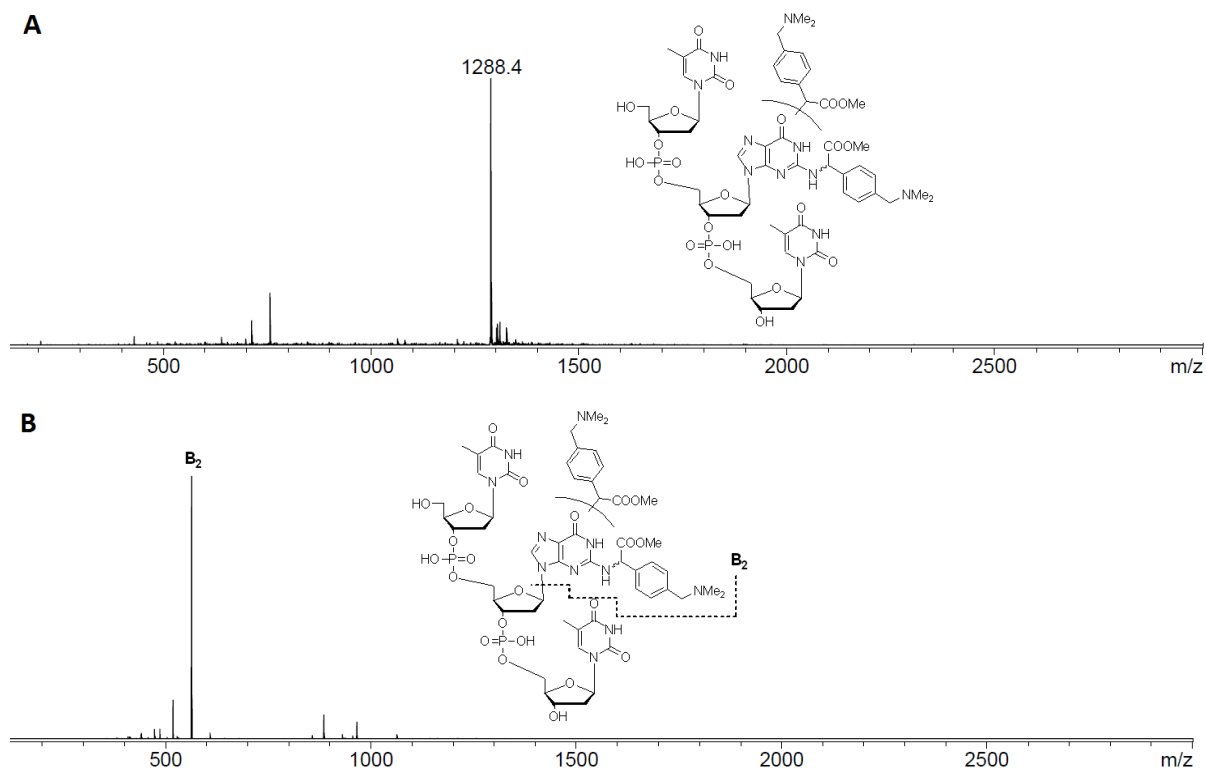
**A**



**B**



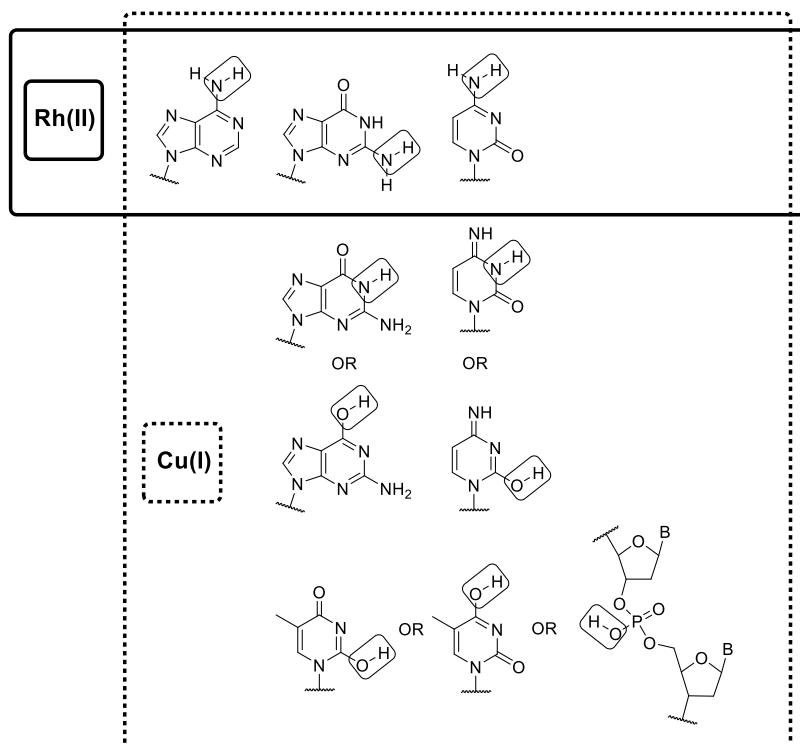
**Figure 18** | HPLC-separation of the reaction mixture from the copper(I)-catalyzed alkylation of d(TCT) (A) and d(TGT) (B) with Dz1. The number for each peak indicates the mass from ESI-MS, the number in parenthesis indicates the number of alkylations per molecule of oligonucleotide.



**Figure 19** | ESI-MS (A) and MS/MS analysis (B) of the doubly-alkylated product of d(TGT) with Dz1 in the presence of copper(I).

RNA also proved to be a viable substrate for carbene insertion (Table 6, Entry 7). Its extensive alkylation producing a number of singly-modified product species clearly demonstrates the capacity of the copper(I)-system for RNA-tagging.

As expected the reactivity of the thymine tetramer was quite modest, resulting in the formation of a number of singly-modified species at an oligonucleotide conversion of 19 % (Table 6, Entry 5). In comparison to the rhodium(II) method, however, such a result demonstrates the significantly higher reactivity of the copper(I) carbenoids, delivering substrate conversion even in the absence of potent nucleophiles on the bases. In this case O-2 and O-4 on the thymine bases or the backbone phosphates could be the potential alkylation targets. Consistent with that observation, modest alkylation yields were obtained with a hair-pin structure with no unpaired reactive bases (Table 6, Entry 8). The observed conversion of 9 % is likely due to targeting nucleotide units in the hair-pin loop. Exposing a reactive base (adenine) as a 3'-overhang (Entry 9) or in the loop region (Entry 10) led to an almost two-fold increase in the alkylation efficiency.



**Scheme 48** | Comparison of the regioselectivity of the rhodium(II) and copper(I) nucleic acid alkylation systems. In addition to the alkylation products observed with the rhodium(II) system several other potential reaction sites are likely to be targeted in the case of copper(I) carbenoid reactions.

The beneficial effect of the THPTA ligand in decreasing the loss of oligonucleotide due to oxidative degradation was once more demonstrated with the recoveries measured for some of the small oligonucleotide substrates (Table 6, Entries 1-5). An optimization of the catalyst in order to further decrease the nucleic acid oxidative damage would bring even more utility to the otherwise quite efficient alkylation system.

### 3.2. Alkylation of simple arylamines in aqueous media

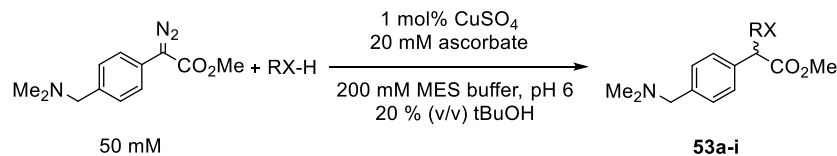
The efficacy of the copper(I)-catalyzed NH-insertion reaction with such complex molecules as DNA and RNA oligonucleotides led to the belief that it would also be of a value for small molecule applications. Moreover, the reaction can be carried out in water which makes it even more attractive considering the development of green technologies.

The N-H insertion with Dz1 and a number of aniline substrates in aqueous media turned out to be remarkably fast and efficient, leading to almost complete conversion of the initial substrates at



reaction times ranging from minutes to a few hours (Table 7, entries 1-9). The individual reactivity of the anilines seemed to have little dependence on their electronic properties (Entries 1, 3-7).

**Table 7** | Scope of copper(I)-catalyzed N-H insertion of Dz1 with nitrogen nucleophiles.



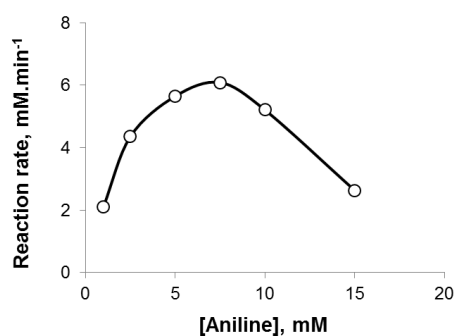
Entry	Product	RX-H	p <i>K</i> <sub>a</sub> <sup>[a]</sup>	Reaction time, h	Conversion, %	Ratio products mono:di <sup>[b]</sup>	Isolated yield monoalkylation, %
1	53a		4.60	0.25	>99	97:3	85
2	53b		4.85	0.25	>99	-	82
3	53c		5.36	1	>99	94:6	75
4	53d		3.99	0.5	>99	98:2	77
5	53e		2.50	0.25	>99	99:1	70
6	53f		2.45	0.25	>99	99:1	74
7	53g		2.76	0.25	>99	99:1	74
8	53h		6.71	20	>99	99:1	51
9	53i		6.03	20	>99	99:1	76
10	-		10.57	24	<2	-	-
11	-		11.05	24	<2	-	-
12	-		11.12	24	<2	-	-
13	-		6.99	24	<2	-	-

[a] p*K*<sub>a</sub> of the singly-protonated cations at 25°C in aqueous solution. [b] For the NH<sub>2</sub>R nucleophiles a second alkylation on the nitrogen is possible. The ratio mono-/dialkylation was determined from the integration areas for the 230-nm absorbance peaks in the UPLC-MS analyses of the crude reaction mixtures.

The catalytically active copper(I) species is easily prepared from a water-soluble copper(II) source, e.g. copper(II)sulfate, chloride or triflate, and sodium ascorbate without the use of any special ligands or additives. That, together with the low catalyst loadings (1 mol%), makes the reaction cost efficient for process applications, where catalyst economy, operational simplicity and the use green solvents, in this case a 4:1 (v/v) mixture of water and *tert*-butanol, are highly desirable. This approach also provides a tool for a strict selection of aromatic over aliphatic amines, as the latter proved to be completely unreactive under the conditions employed (Table 2, entries 10-12). This could be explained with the protonation state of the N-H group, as at pH 6 the three aliphatic amines tested are almost entirely protonated due to the higher  $pK_a$  values of the conjugate acids, thus a lot less nucleophilic. Several important features of the copper(I) system were discovered upon a more rigorous investigation of the reaction mixtures. First, all primary arylamines delivered a small amount of doubly-alkylated product, as more electron-rich anilines seemed to favor its formation (Table 7). Second, small amounts of O-H insertion products with water and ascorbate were found. The possibility for targeting alcohols was further examined with 1-butanol, ethylene glycol and phenol, but no insertion products were detected for any of these substrates. In that sense the reactivity of the water can be explained with its large excess in the reaction system. In the case of ascorbate the quite acidic O-H group at C-3 ( $pK_a$  4.25) is to be considered as a potential reaction site.<sup>94</sup>

Strongly coordinating Lewis basic substrates (imidazole (Table 7, Entry 13), 2-(diethylamino)ethanethiol) completely blocked the catalytic activity of copper(I), as neither formation of the respective X-H insertion products, nor products from the side-reactions of O-H insertion in water and ascorbate were detected.

Establishing a mechanistic model for the functioning of the copper(I) cabenoid insertion proved to be complicated even for the reaction with small anilines. The main problem was the presence of multiple coordination states of the catalytically active copper(I) species. The copper(I) complex should have a tetrahedral geometry with four coordinating ligands. In the present reaction system water, ascorbate, THPTA, as well as the diazo substrate can bind to the metal center. In addition, attempts to correlate the initial rates of Dz1 consumption with the aniline concentration in a reaction of N-H insertion did not show a Michaelis-Menten type saturation kinetics. Instead a clear effect of substrate inhibition was observed at concentrations above 10 mM, which indicated coordination of the aniline substrate to the copper catalyst (Figure 20).



**Figure 20** | Dependence of the initial rate of Dz1 conversion on the aniline concentration in a Cu(I)-catalyzed N-H insertion in aniline. Reaction conditions: 25  $\mu$ M CuSO<sub>4</sub>, 12.5 mM Dz1, 20 mM sodium ascorbate, 200 mM MES buffer, pH 6.0.

### 3.3. Tandem auto-catalytic Cu(I)-catalyzed azide-alkyne cycloaddition/N-H insertion. Biomolecular labeling applications

Apart from its efficiency of aniline alkylation with carbenoids, a great advantage of the copper(I) catalytic system is the possibility for its multi-purpose use. The active catalytic species could be utilized for carrying out several concurrent transformations in a more economic one-pot set-up, achieving a greater product diversity with just a single purification step.

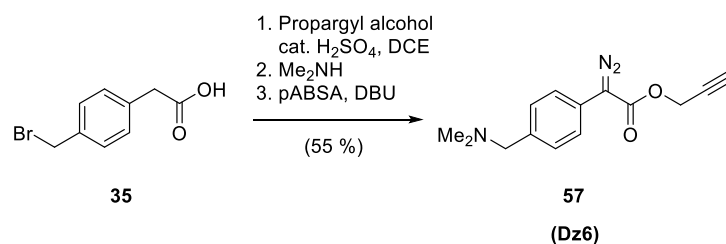
**Table 8** | Auto-tandem catalytic CuAAC/N-H insertion with small molecule substrates

Entry	Amine	Diazo compound	Azide	Product	Isolated yield, %
1					62
2					70

Reaction conditions: 0.5 mM CuSO<sub>4</sub>, 50 mM amine, 50 mM diazo substrate, 55 mM azide, 35 % (v/v) tBuOH, 24 h, room temperature.

The utility of the CuAAC in both chemical library construction and macromolecular conjugation makes it an ideal candidate for complementing the N-H insertion protocol in a concurrent reaction set-up. A tandem copper(I)-catalyzed N-H insertion/CuAAC was first demonstrated on small molecule substrates (Table 8). The two examples shown differ by the degree of complexity of the azide moiety and the location of the alkyne group. The tandem processes worked well in both cases with very good yields of the isolated ‘click’/insertion products **55** and **56**. Since copper(I) tends to form rather stable complexes with alkynes, it is important to use a slight excess of the azide component in order to insure complete conversion of the alkyne and make the catalyst available for the insertion reaction.

The ultimate task at hand was achieving nucleic acid tagging through copper(I)-catalyzed tandem ‘click’/N-H insertion. In order to do that the alkyne moiety for the copper ‘click’ reaction was installed on the diazo compound, which in this case would be the connecting element between the oligonucleotide and the appropriately functionalized azide. A propargyl ester analogue of Dz1 was prepared as shown on Scheme 49. Esterification of the commercially available 4-(bromomethyl)phenylacetic acid with a small excess of propargyl alcohol according to the modified protocol of Clinton and Laskowski,<sup>95</sup> followed by substitution of the benzylic bromide with dimethyl amine and installation of the diazo group via nucleophilic diazo transfer afforded the final compound Dz6 (**57**) in a 55 % yield over three steps.

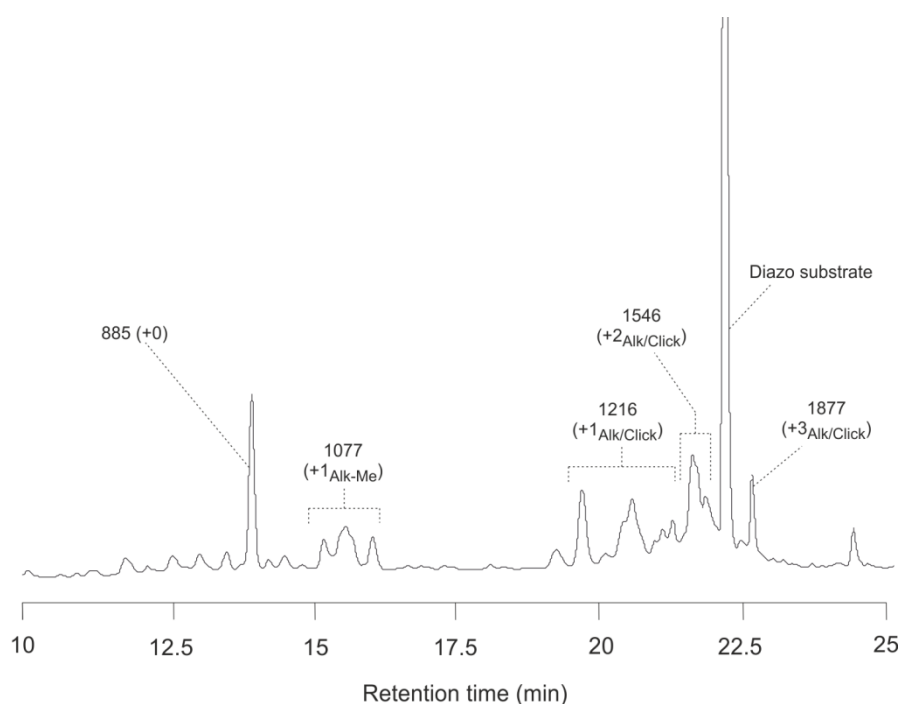


**Scheme 49** | Synthetic scheme for preparation of Dz6 (**57**).

The tandem CuAAC/N-H insertion with Dz6 and the two azides **47** and **54** was tested on three short DNA oligonucleotides, using the reaction conditions for the nucleic acid alkylation with copper(I)-carbenoids. The CuAAC reaction appeared to be quite efficient and quick, as in all cases the modification products detected also carried the ‘click’ tag.

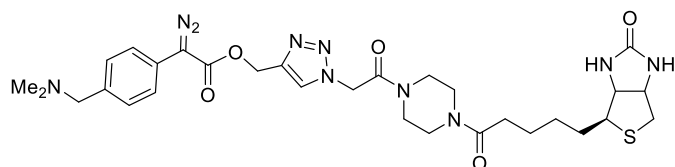
The modification of d(ATG) with Dz6 and 3-azidopropan-1-ol yielded a number of singly-, doubly-, and triply-tagged products at a conversion of 80 % after 18 h (Figure 21). Tandem MS analysis showed the presence of ‘click’/alkylated guanine base in all product fractions and modified adenine only in the triply-alkylated species. This indicates that guanine was the primary alkylation target, undergoing two consecutive modifications before adenine could be targeted. In addition to that some singly-alkylated products with a hydrolyzed ester moiety were observed.

The G-rich sequence d(GGA GGC) was modified as described for d(ATG). An RP-HPLC-analysis of the reaction mixture under the standardly used neutral conditions yielded quite a poor peak shape and separation (Figure 22). Such deviations in the chromatographic behavior had been previously observed with other G-rich sequences and were most likely due to intermolecular association effects. The modification itself was quite efficient, as a number of product species with up to four ‘click’/N-H insertion groups per molecule of oligonucleotide were detected. An oligonucleotide conversion of roughly 60 % was determined on the base of the peak area, although a precise estimate could not be made due to the presence of non-modified starting oligonucleotide in all fractions (Figure 23).



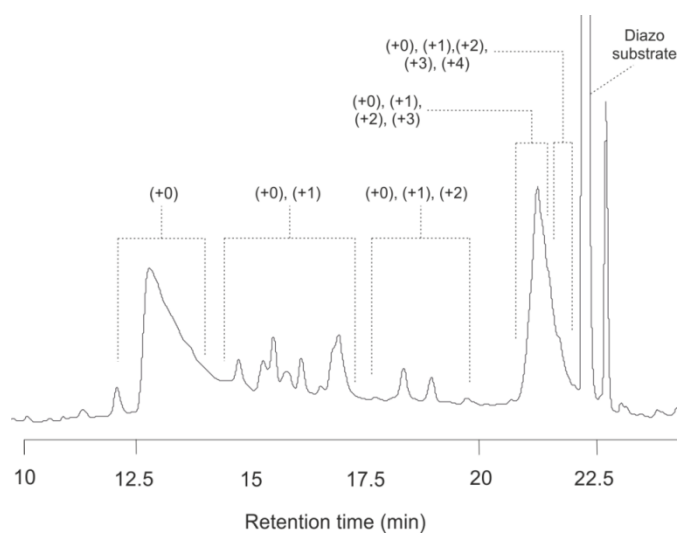
**Figure 21** | HPLC analysis of the reaction mixture of the tandem CuAAC/N-H insertion labeling of d(ATG) with Dz6 and 3-azidoprop-1-ol. The fractions were identified by ESI-MS analysis (the masses are given for each peak). The number and type of modification is indicated in parenthesis. Alk-Me: singly-modified species with hydrolyzed ester group (only acyl portion remaining).

The modification of d(ATGC) with Dz8 and biotin azide **47** resulted in a lot cleaner product mixture with up to two ‘click’/N-H insertion modifications per oligonucleotide molecule detected (Figure 24). The overall oligonucleotide conversion was lower than for the example with d(ATG). As pointed out, CuAAC between Dz6 and biotin azide **47** should be the first reaction to take place and should result in the *in situ* formation of the diazo compound **57**.

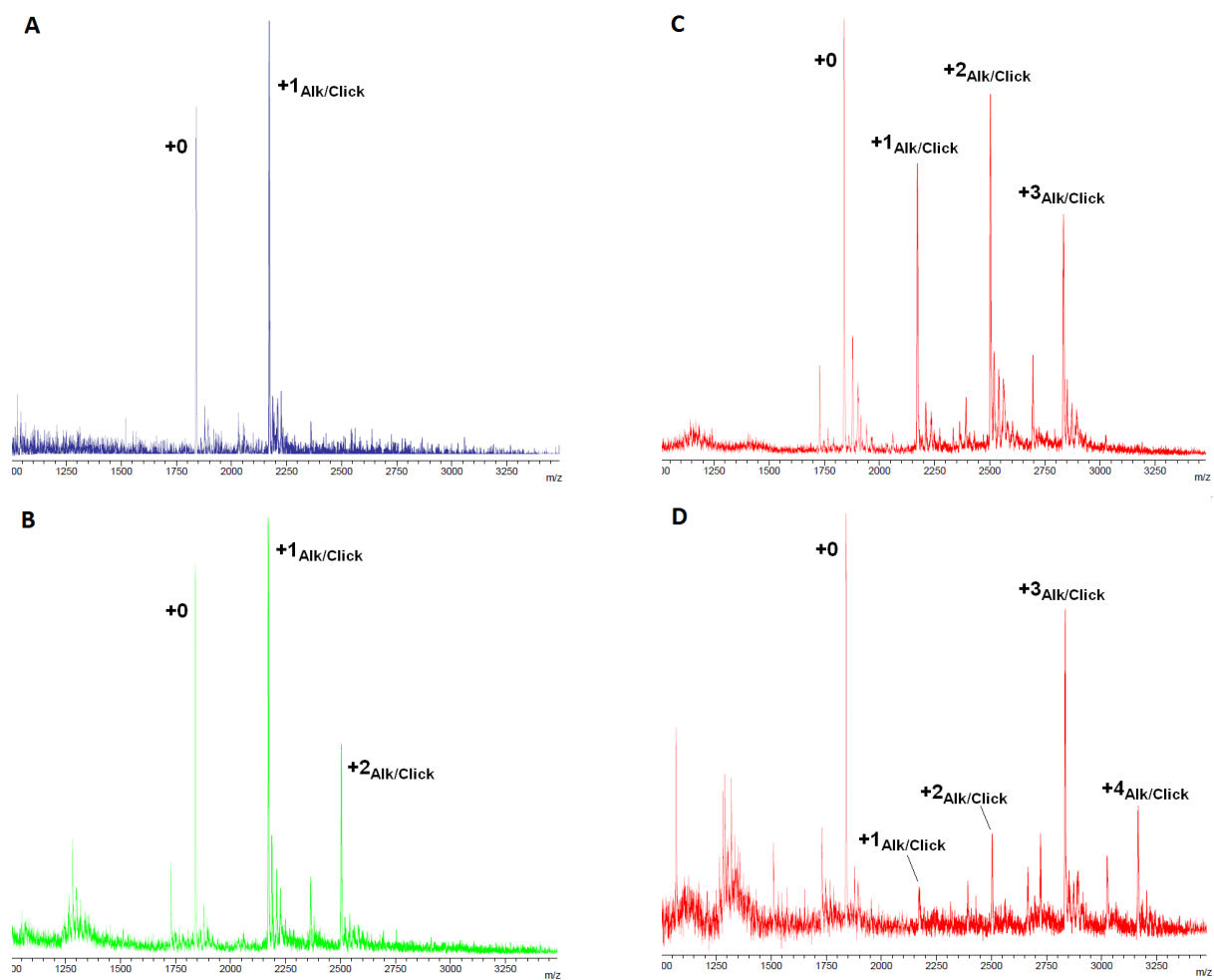


57

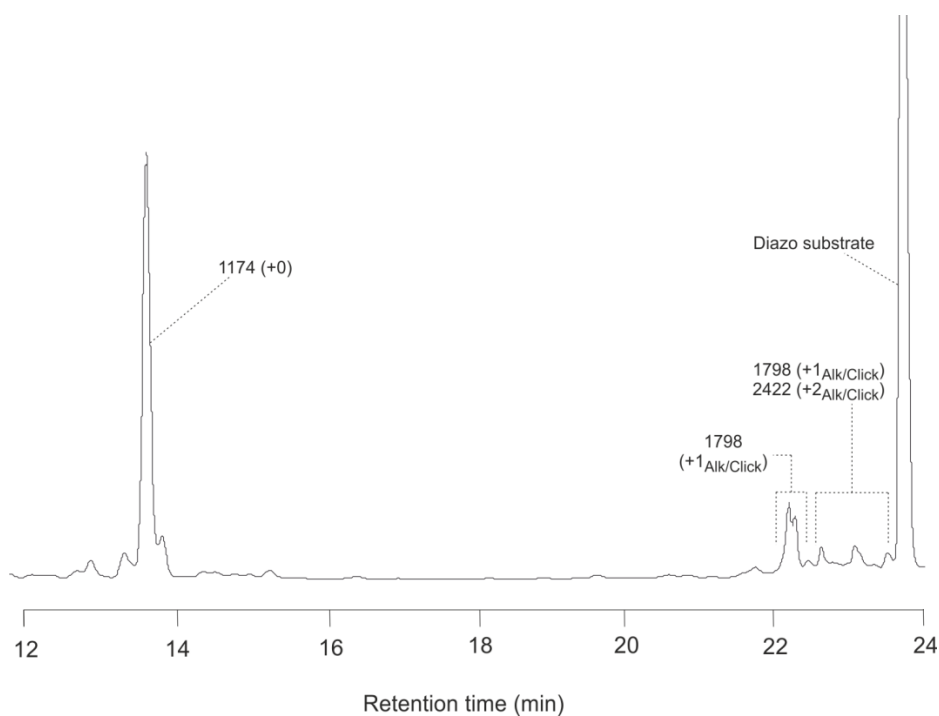
Subsequent DNA modification with the respective copper(I)-stabilized carbenoid was most likely hindered by the sterically demanding alkyl ester portion. The increased steric bulk was also likely the reason for the lack of detectable amounts of products with higher modification numbers.



**Figure 22** | HPLC analysis of the reaction mixture of the tandem CuAAC/N-H insertion labeling of d(GGA GGC) with Dz6 and 3-azidoprop-1-ol. The fractions were identified by MALDI-TOF MS. The number and type of modification is indicated in parenthesis. Alk-Me: singly-modified species with hydrolyzed ester group (only acyl portion remaining).



**Figure 23** | MALDI TOF of the products from the tandem 'click'/alkylation reaction of d(GGA GGC) with Dz6 and 3-azidoprop-1-ol. The number and type of modification is indicated in parenthesis.



**Figure 24** | HPLC analysis of the reaction mixture of the tandem CuAAC/N-H insertion labeling of d(ATGC) with Dz6 and biotin azide **47**. The fractions were identified by MALDI-TOF MS (the masses are given for each peak). The number of modifications are indicated in parenthesis.

An interesting point with respect to the potential labeling applications of the method is its ability to target proteins or, more importantly, the ability to modify nucleic acids in the presence of proteins (e.g. in cell lysates). This was tested with three proteins, BSA, lysozyme and wild-type streptavidin, which were present at a 1 mM concentration in a standard d(TAT) alkylation reaction with Dz1 as described in Section 3.1. Lysozyme and streptavidin did not significantly affect the oligonucleotide alkylation efficiency or the overall catalyst reactivity as judged from the conversions of d(TAT) and the Dz1 (Table 9). The presence of BSA on the other hand significantly decreased the catalytic activity and almost completely blocked the oligonucleotide alkylation reaction. This means that although suitable for some *in vitro* applications the method has to be examined in terms of feasibility on a case-to-case basis to exclude possible interference of other components of the system.

**Table 9** | Cu(I)-Catalyzed alkylation of d(TAT) with Dz1 in the presence proteins

Entry	Protein	Conversion of d(TAT), %	Conversion of Dz1, %
1	-	41	89
2	Bovine serum albumin	<1	32
3	Lysozyme	37	91
4	Streptavidin (wild type)	33	82



### 3.4. Conclusions and future directions

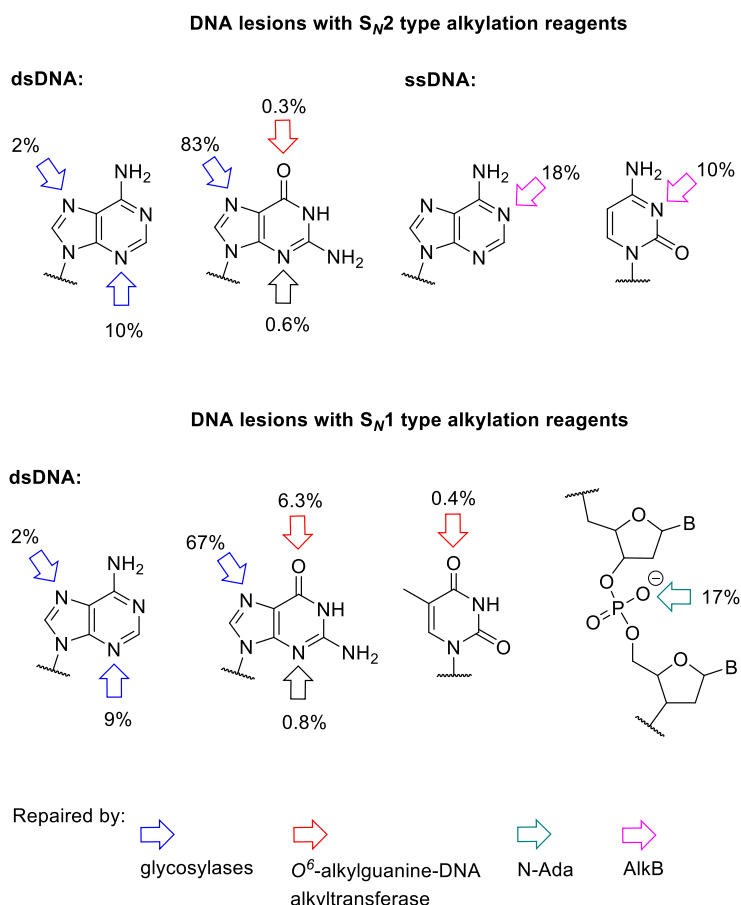
A catalytic system for arylamine alkylation based on N-H insertion of copper(I)-stabilized carbenoids, has been developed. The catalyst is conveniently prepared *in situ* from a copper(II) salt and a reducing reagent and utilized under the standard conditions for CuAAC. A wide scope of arylamine substrates ranging from simple anilines to nucleobases on short DNAs and RNAs has been efficiently targeted, achieving high substrate conversions. The enhanced reactivity of the copper(I) carbenoids towards nucleic acid substrates allows for multiple targeting, especially in guanine-containing substrates, making the method suitable for biomolecular tagging. On the other hand the efficiency with anilines in aqueous media and the strict selection for aromatic over aliphatic amines suggests potential applications in synthetic chemistry. The complete convergence of the reaction conditions with those for CuAAC allowed copper ‘click’ chemistry to be combined with copper(I) carbenoid N-H insertion in an auto-catalytic tandem process. Such a convenient one-pot set-up provides direct access to a number of diverse structures for small molecule library applications requiring only a single purification step at the end. It could also be used for direct DNA and RNA tagging with an alkyne-bearing diazo substrate as a linker unit for coupling the nucleic acid with a certain azide-modified marker.

A continuation of this study should address several questions of potential fundamental and practical value:

- (a) Determination of the regioselectivity of nucleobase targeting – revealing of the position of the second alkylation in the case of guanine and cytosine;
- (b) Copper(I) coordination states - formation and catalytic activity of Cu(I)/nucleic acid and Cu(I)/aniline complexes;
- (c) Reactivity of copper(I) carbenoids on longer DNAs and RNAs. As with the rhodium(II) catalytic system this would require analytical methods of higher sensitivity and precision;
- (d) Development of an asymmetric version of the aniline alkylation system – screen and design of chiral copper(I) ligands for use under aqueous conditions;
- (e) Implementation of the system in the context of a sequence-specific nucleic acid alkylation. Combined with a further optimization of the reaction conditions in order to suppress the nucleic acid oxidative damage, such a system would be of a great utility for some complex chemical biology applications.

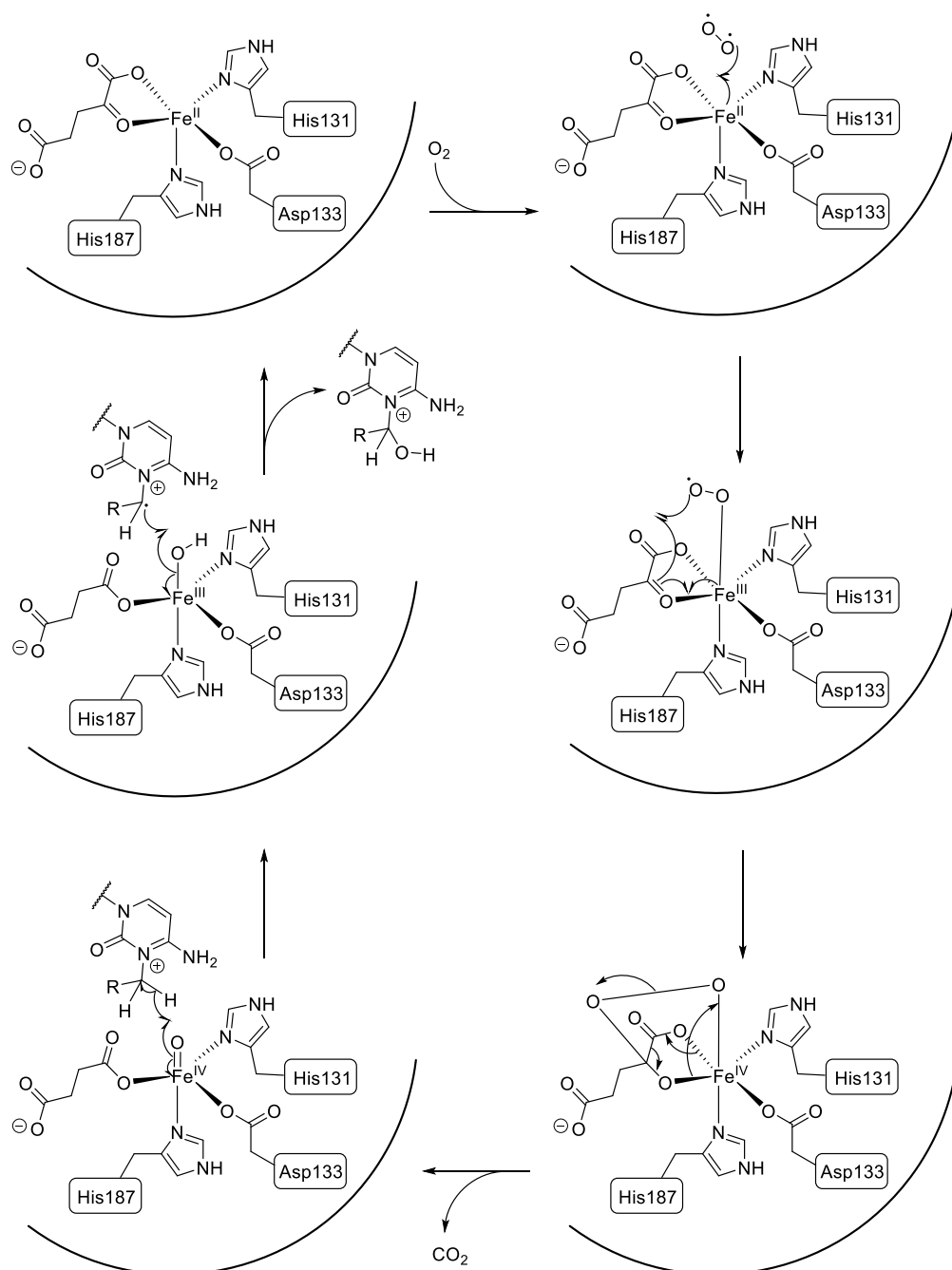
## Chapter 4. Direct reversal of DNA alkylation with AlkB dioxygenase from *E.coli*

Cell DNA is constantly subjected to modifications by both endogenous and exogenous agents. In some cases these modifications are important for essential processes such as regulation of the DNA replication and transcription. In other cases, especially for reagents coming from the environment, such nucleic acid modifications could be toxic or mutagenic. In this sense DNA repair is an essential mechanism for insuring proper cell functioning. One such type of DNA repair is the direct repair, in most cases known this is the direct reversal of DNA alkylation damage. There are several proteins known to perform such a direct alkylation reversal: the N-Ada protein in *E.coli*, the O<sup>6</sup>-alkylguanine-DNA alkyltransferase family and the AlkB family. An outline of the most common types of DNA alkylation damage and the respective enzymes for their direct repair is shown in Scheme 50.<sup>37</sup>



**Scheme 50** | Typical types of DNA damage observed with S<sub>N</sub>2 and S<sub>N</sub>1 alkylating reagents. The fraction of the particular alkylation damage together with the respective enzyme for direct repair are indicated for each site in single- and double-stranded DNAs.

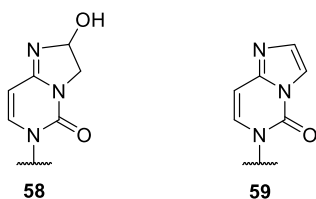
The present study focuses on DNA dealkylation catalyzed by AlkB. AlkB is an iron(II)/ $\alpha$ -ketoglutarate dependent dioxygenase, which is responsible for removal of alkylations mainly from N-1 of adenine and N-3 of cytosine. A plausible mechanism for its catalytic action is shown in Scheme 51. The iron(II) species involved in the catalysis is directly ligated to His131, Asp133 and His187 of the enzyme active site, as well as to an  $\alpha$ -ketoglutarate anion.



**Scheme 51** | Plausible mechanism for the direct repair of N-alkylated DNA nucleobases by the dioxygenase AlkB from *E. coli*. The mechanism shown is for dealkylation of cytosine base. The N<sup>3</sup>-hydroxyalkylcytosine breaks down further to yield the respective aldehyde and the unmodified nucleobase.

The catalytic cycle starts with coordination of molecular oxygen to the iron(II) center to form an iron(III)-bound superoxide anion radical. The superoxide anion radical then forms a bridged peroxo-type intermediate with the C-2 of the  $\alpha$ -ketoglutarate and the iron center (now oxidized to iron(IV)). Decarboxylation of this intermediate accompanied by a polar cleavage of O-O-bridge gives a succinate anion and the key iron(IV)-oxo intermediate. This species abstracts a hydrogen radical from the nucleic acid substrate  $\alpha$ -N-alkyl carbon to give an N-alkyl radical and an iron(III)-hydroxyl species. A subsequent hydroxylation of the N-alkyl radical yields an N-hydroxyalkyl nucleobase, regenerating the iron(II) center. The N-hydroxyalkyl derivative breaks down to an alkyl aldehyde and the unmodified base. Several different elements of the nucleic acid substrate structure proved to be important for the dealkylation reaction. The requirement for a 5'-phosphate which most likely assists the proper positioning of the substrate is the first one.<sup>96</sup> The smallest efficiently dealkylated substrate was found to be 1-methyl-2'-deoxyadenosine-5'-monophosphate, while the respective 3'-monophosphate and 5'-triphosphate were quite poorly converted mainly due to impaired binding. Concerning the nucleobase alkylations at N-1 of adenine and N-3 of cytosine are the primary targets, although N-3 in thymine and N-1 in guanine could also be repaired.<sup>97</sup>

So far damage reversal by AlkB has been shown only for N-methyl derivatives of the four canonical DNA bases, as well as for N<sup>3</sup>-ethyl cytosine<sup>97</sup> and two bicyclic cytosine derivatives, 3,N<sup>4</sup>- $\alpha$ -hydroxyethanocytosine (**58**) and 3,N<sup>4</sup>-ethenocytosine (**59**).<sup>98</sup> Studies on the repair of propyl, hydroxyethyl and hydroxypropyl modifications have only been done *in vivo* with poorly defined nucleic acid substrates.<sup>96</sup>



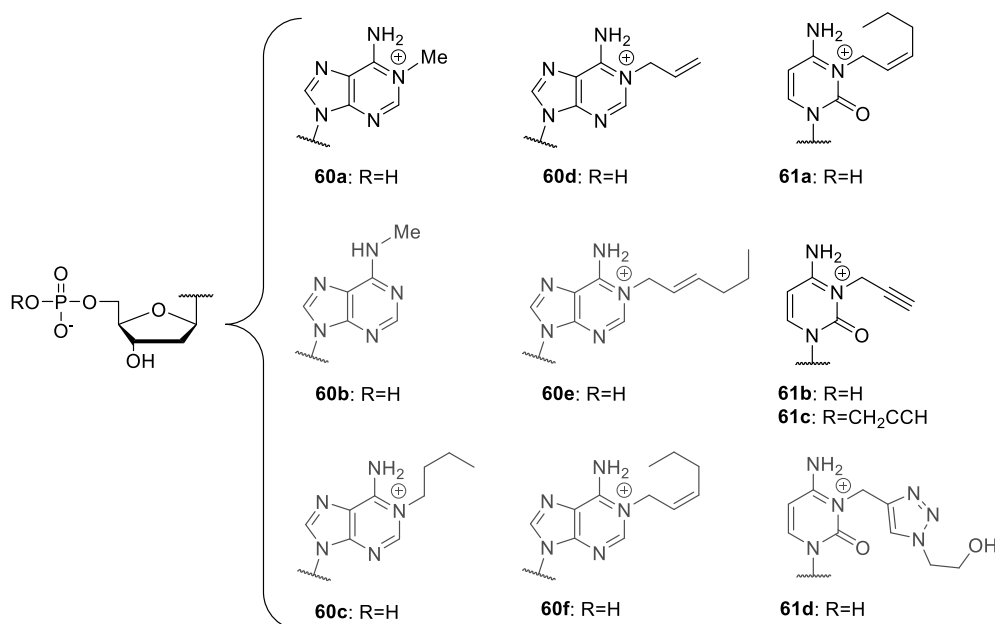
The objective of the work in this chapter is to extend the variety of N-alkyl groups on 2'-deoxynucleoside substrates repaired by AlkB from *E.coli* and to optimize the substrate structure in order to increase the efficacy of repair. Such system could then be used for *in vitro* and *in vivo* removal of previously installed alkyl moieties, bearing various tags and markers (on-demand tag removal). Our test concept was that if the cargo was attached to the DNA through a medium-sized or a long alkyl group spacer it should remain outside the catalytic pocket of the enzyme, thus its steric bulk would have less interference on the repair process.

#### 4.1. Direct alkylation of mononucleoside-5'-monophosphates

N<sup>1</sup>-adenine and N<sup>3</sup>-cytosine alkylated substrates for the reversal repair study were prepared by treatment of the corresponding 2'-deoxynucleoside 5'-monophosphates with alkyl bromides or iodides as already described for N<sup>1</sup>-isopentenyladenosine.<sup>99</sup> The reactions were carried out in DMSO at room temperature varying the amount of alkylation reagent and the reaction time. The procedure worked well only for the adenine N-1 alkylation with methyl iodide, where a pure product in a very good yield (80 %) was isolated already after the initial work up. In all other cases less than optimal conversions (<50 %) were observed and attempts to improve them by increasing the amount of alkylation reagent led to the formation of doubly-modified products by targeting the 5'-phosphate. The alkylation was more efficient for the conjugated alkyl bromides (allyl, propargyl, *cis*- and *trans*-2-hexenyl) and shorter reaction times were required to achieve moderate substrate conversions. The rather complicated purification, however, resulted in isolated yields of the target product of not more than 10 %.

#### 4.2. Expanding the substrate scope of AlkB: direct enzymatic removal of bulky N-alkyl groups from nucleoside-5'-monophosphates

The direct dealkylation catalyzed by AlkB was tested on a small series of adenine and cytosine derivatives with varying length and structure of the N-alkyl group (Scheme 52) under conditions similar to those described earlier.<sup>100</sup>



**Scheme 52** | Scope of N-alkylated 2'-deoxyadenine and 2'-deoxycytidine 5'-monophosphates tested in a dealkylation reaction with AlkB.

Previous kinetic studies on catalytic repair of several small N-methylated adenine-containing derivatives with AlkB showed  $K_M$ -values of up to 50  $\mu\text{M}$  even for the poorest substrates.<sup>96</sup> Based on that the activity measurements were carried out with 1 mM initial concentration of the alkylated substrate to insure a zero-order behavior of the system with respect to it. The obtained reaction rate values were compared to that for the reference substrate **60a** under equal other conditions.

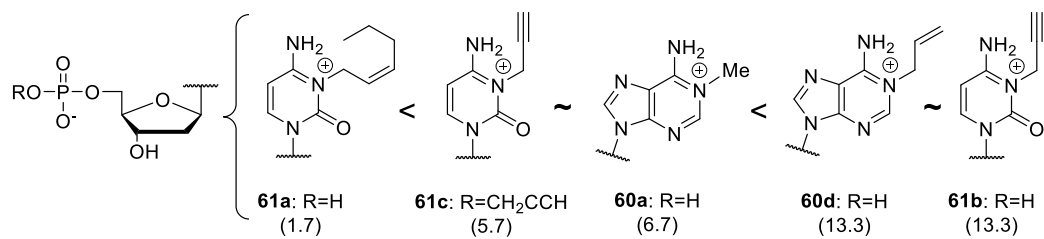
The simple increase in the length of the N<sup>1</sup>-alkyl chain as in the case of N<sup>1</sup>-*n*-butyladenine resulted in poor dealkylation rates with less than 10 % overall substrate conversion after 24 h incubation with the enzyme. The enzyme was also unable to repair alkylation at N-6 of adenine as demonstrated in experiments with compound **60b**.

Since the mechanism of dealkylation most likely involves a radical hydroxylation of the N-alkyl  $\alpha$ -carbon (Scheme 51), alkyl groups that could stabilize a radical at that position should be removed more efficiently. Indeed, tests with the N<sup>1</sup>-allyl-2'-deoxyadenosine derivative **60d** showed that its AlkB-catalyzed repair was two-fold faster than that of the reference substrate **60a**. Consistent with that the N<sup>3</sup>-propargyl group in the cytidine derivatives **61b** and **61c** was also readily removed. In the same time the 5'-O-alkylation in **61c** had a negative effect on the dealkylation, which was more than two-fold slower than that for compound **61b**.

The possibility to exploit the 2,3-double bond motif in both *cis*- and *trans*-configuration for increasing the efficiency of removal of longer alkyl chains was tested with adenine derivatives **60e** and **60f**. Both substrates displayed quite a slow dealkylation delivering less than 10 % conversion after 24 h incubation with the enzyme. On the other hand the N<sup>3</sup>-(*cis*-2-hexenyl)-2'-deoxycytidine derivative **61a** showed significantly higher propensity for enzyme repair, as 10 % substrate conversion was achieved in less than an hour reaction time. This might be a result of decreasing the base size, thus facilitating the substrate binding in the enzyme active site.

Finally, a dealkylation reaction was attempted on the sterically demanding substrate **61d** bearing a triazole-containing N<sup>3</sup>-alkyl moiety. The increased bulk in a close proximity of the nucleobase was most likely the reason for the lack of detectable dealkylation of this substrate. This result once more confirms the need for a linker in case functionalization with larger cargo molecules is desired.

An activity comparison of the best substrates for AlkB is outlined in Scheme 53.



**Scheme 53** | Scope of the N-modified 2'-deoxyadenine and 2'-deoxycytidine 5'-monophosphates dealkylated by AlkB. Reaction rates from the activity assays [ $\mu\text{M}\cdot\text{min}^{-1}$ ] are given in parenthesis. Assay conditions: 1 mM alkylated nucleoside monophosphate, 2 mM  $\alpha$ -ketoglutarate, 100  $\mu\text{M}$  FeSO<sub>4</sub>, 2 mM sodium ascorbate, 2 mM DTT, 50 mM KCl, 100  $\mu\text{g}\cdot\text{ml}^{-1}$  BSA, 50 mM Tris-HCl, pH 8.0, 7.44  $\mu\text{M}$  AlkB, 37°C.

### 4.3. Conclusions and future directions

The scope of N-alkylated adenine and cytosine substrates for the dioxygenase AlkB from *E.coli* has been expanded, as the enzyme was shown to efficiently remove small and medium-sized conjugated N-alkyl moieties. Such direct repair is likely favored by the ability of such functionalities to stabilize carbon radicals, thus facilitating the radical hydroxylation step of the enzyme reaction. The possibility larger tags connected to a nucleobase of the DNA substrate through a suitable linker to be enzymatically removed is demonstrated with the dealkylation of N<sup>3</sup>-(cis-2-hexenyl)-2'-deoxycytidine-5'-monophosphate. However, some more detailed studies are needed to fully assess the system properties and define more precisely its possible practical applications:

- (a) Optimization of the activity assay to improve the enzyme stability – concentration of the iron(II) source, use of TCEP instead of DTT;
- (b) Activity tests with an optimized version of the N<sup>3</sup>-alkylated 2'-deoxycytidine-5'-monophosphate substrate: dealkylation efficiency with a C5-C7 alkyl chain substrates containing a 2,3-alkyne group; enzymatic removal of the same functionalities with an attached cargo (rhodamine, biotin, etc.);
- (c) Determination of the steady-state kinetic parameters for the AlkB-catalyzed dealkylation of substrates **60d** and **61b**: separation of the kinetic constants  $k_{cat}$  and  $K_M$  would clarify if the respective C3-moieties improve the substrate binding or the catalytic step of the enzymatic reaction. An increase in the turnover number  $k_{cat}$  would be an indirect evidence for the involvement of a radical species in the catalytic process as suggested by the widely accepted mechanism for the enzyme action;
- (d) Enzyme activity tests with short oligonucleotides of the type d(TTCTT), where cytosine is N<sup>3</sup>-alkylated – dependence of the dealkylation efficacy on the substrate base length.
- (e) Enzymatic repair of N<sup>1</sup>- and N<sup>2</sup>-alkylated guanine derivatives.



## Chapter 5. Experimental

### 5.1. General

All reagents and solvents used were of analytical grade. Buffers were prepared with ultrapure water. All chemicals were purchased from Sigma-Aldrich, Fluka or Acros and were used as received. Analytical TLC was performed on Silica gel 60 F<sub>254</sub> pre-coated aluminium sheets. Flash chromatography was performed on Silica gel 60 40-63  $\mu\text{m}$  (230-400 mesh) (SiliCycle, Quebec).

Solid-phase oligonucleotide synthesis was carried out on an Expedite 8909 nucleic acid synthesis system (PerCeptive Biosystems). Phosphoramidites for DNA synthesis ( $\text{A}^{\text{Bz}}$ ,  $\text{G}^{\text{dmf}}$ ,  $\text{C}^{\text{Ac}}$ , T) were purchased from Proligo Reagents, SAFC. SynBase 500/110 AllFit pre-modified oligonucleotide synthesis columns, Acetonitrile anhydrous wash, BTT activator, CapA, CapB and Oxidizer for the Expedite device, as well as 2'-O-TBDMS-protected RNA phosphoramidites ( $\text{A}^{\text{Bz}}$ ,  $\text{G}^{\text{dmf}}$ ,  $\text{C}^{\text{Ac}}$ , T) were purchased from Link Technologies.

$^1\text{H}$ ,  $^{13}\text{C}$  and  $^{31}\text{P}$  NMR spectra were acquired on a Bruker Avance III 600 MHz, Bruker Avance III+ 500 MHz and Bruker Avance III 400 MHz spectrometers at 298 K.  $^{19}\text{F}$  NMR spectra were measured on a Bruker Avance III 400 MHz spectrometer at 298 K. Chemical shifts relative to TMS were referenced to the solvent's residual peak and are reported in ppm.

ESI MS spectra were obtained on a Bruker Esquire3000plus spectrometer by direct injection in positive polarity of the ion trap detector. Tandem MS spectra were measured on the same device after collision-induced dissociation with molecular nitrogen. High resolution mass spectra were acquired on a Bruker maXis 4G QTOF ESI mass-spectrometer. EI MS were obtained on a Finnigan MAT 95 device. MALDI TOF analyses were carried out on a Bruker Microflex mass spectrometer in linear negative mode using the following matrixes: 3-hydroxypicolinic acid/ammonium citrate buffer, pH 6.0 (for oligonucleotides with  $M_r > 3000$ ) or 2,4,6-trihydroxyacetophenone/ammonium citrate buffer, pH 6.0 (for oligonucleotides with  $M_r < 3000$ ). IR spectra were obtained on a Bruker ALPHA FT-IR spectrometer.

## 5.2. HPLC

Analytical HPLC analyses were carried out on an Agilent 1100 LC system equipped with a Chromolith Performance RP18e 4.6 × 100 mm column (Merck) using methods A, B, C or D or with an Eclipse XDB-C18 4.6 × 150 mm column (Agilent), using method E or F.

*Method A:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 7.1)/acetonitrile gradient 0-16 % acetonitrile in 12 min, 16-80 % acetonitrile in 3 min, 80 % acetonitrile in 2 min.

*Method B:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 7.1)/acetonitrile gradient 0-16 % acetonitrile in 12 min, 16-80 % acetonitrile in 15 min, 80 % acetonitrile in 2 min.

*Method C:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 7.3)/acetonitrile gradient 0-16 % acetonitrile in 12 min, 16-80 % acetonitrile in 3 min, 80 % acetonitrile in 2 min.

*Method D:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 6.0)/acetonitrile gradient as in method A.

*Method E:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 7.3)/acetonitrile gradient 0-16 % acetonitrile in 18 min, 16-80 % acetonitrile in 4.5 min, 80 % acetonitrile in 3 min.

*Method F:* Flow rate 1 ml/min, 100 mM triethylammonium acetate (pH 7.1)/acetonitrile gradient 0-3 % acetonitrile in 4 min, 3-80 % acetonitrile in 4 min, 80 % acetonitrile in 1 min.

In all methods detection was done by monitoring the absorbance of the column effluent at 254 nm.

UPLC-MS was carried out on a Agilent 1290 Infinity system, equipped with a Zorbax Eclipse Plus C18 2.1 × 50 mm column (Agilent), coupled to an Agilent 6130 Quadrupole LC/MS. Elution was done with 0.1 % trifluoroacetic acid in water/acetonitrile system at a flow rate of 0.2 ml/min using method D: 0-54 % acetonitrile in 4 min, 54-95 % acetonitrile in 1 min, 95 % acetonitrile in 1 min, detection at 230 and 254 nm, and with ESI-MS in positive ion mode of the ion trap.

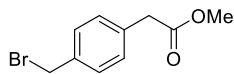
Semi-preparative HPLC was carried out on a Shimadzu VP Liquid Chromatograph using a LiChroCART 10 × 250 mm cartridge, packed with Purosphere®STAR RP-18e 5 µm (Merck). Elution was done at a flow rate of 5 ml/min in a 100 mM triethylammonium acetate (pH 7.1)/acetonitrile gradient 0-16 % acetonitrile in 15 min, 16-80 % acetonitrile in 20 min, 80 % acetonitrile in 2 min. Detection was done by monitoring the absorbance of the column effluent at 254 nm.

Preparative HPLC was carried out on a Shimadzu Prominence Preparative Liquid Chromatograph using a Gemini C18 10 µm 21.2 × 250 mm column (Phenomenex). Elution was done at a flow rate of 20 ml/min in water/acetonitrile gradient 0 % acetonitrile in 3 min, 0-40 % acetonitrile in 27 min, 40-95 % acetonitrile in 5 min, 95 % acetonitrile in 5 min. Detection was done by monitoring the absorbance of the column effluent at 254 nm.

### 5.3. Chemical synthesis

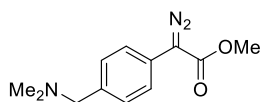
Compound **11** was prepared as described by Antos and Francis,<sup>54</sup> THPTA – as described by Finn et al.<sup>86</sup>

**Methyl 2-(4-(bromomethyl)phenyl)acetate (36).** 4-(Bromomethyl)phenylacetic acid (0.2 g, 0.87 mmol) was mixed with dry methanol (2 ml) and dry benzene (2 ml) under nitrogen. Trimethylsilyldiazomethane (2 mol/l in hexanes) was added drop-wise



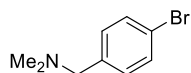
at continuous stirring until stable yellow color of the solution was obtained (the reaction vessel should be vented as gas evolution takes place). The mixture was stirred for 15 min at room temperature, then evaporated and dried under high vacuum to yield the pure compound as a colorless oil (0.21 g, 0.86 mmol, quant.). TLC (n-hexane:Et<sub>2</sub>O 4:1 v/v) R<sub>f</sub>=0.36. <sup>1</sup>H NMR (500.1 MHz, CDCl<sub>3</sub>) δ, ppm: 7.36 (d, J = 8.1 Hz, 2H), 7.26 (d, J = 8.2 Hz, 2H), 4.48 (s, 2H), 3.69 (s, 3H), 3.62 (s, 2H). <sup>13</sup>C NMR (125.8 MHz, CDCl<sub>3</sub>) δ, ppm: 171.73, 136.68, 134.28, 129.75, 129.31, 52.16, 40.86, 33.21. MS (EI): 104.1 [M-Br-COOMe]<sup>+</sup>, 163.1 [M-Br]<sup>+</sup>, 183.0, 185.0 [M-COOMe]<sup>+</sup>, 242.0, 244.0 [M]<sup>+</sup>.

**Methyl 2-diazo-2-(4-((dimethylamino)methyl)phenyl)acetate (37, Dz1).** Compound **36** (0.2 g, 0.82 mmol) was dissolved in 2 ml of absolute ethanol, dimethylamine in absolute ethanol (2 ml, 5.6 mol/l) was added and the mixture was stirred at room temperature for 30 min. The volatiles were then removed under vacuum and



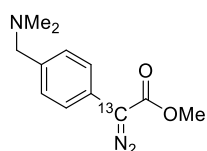
the residue was partitioned between ethyl acetate and saturated sodium bicarbonate solution. The organic phase was washed with brine, dried with anhydrous sodium sulfate and evaporated under vacuum. The remaining yellow oil was dried under high vacuum, and then mixed with p-acetamidobenzenesulfonyl azide (pABSA, 0.216 g, 0.90 mmol) in 5 ml of dry acetonitrile under nitrogen. DBU (0.187 g, 1.23 mmol) was added to the solution with continuous stirring and the mixture was stirred for 24 h at room temperature. The mixture was then evaporated under vacuum and the residue purified by flash chromatography on Si60 in DCM/methanol to afford a yellow-orange oil (0.082 g, 0.35 mmol, 43 %). TLC (DCM:methanol 10:1 v/v) R<sub>f</sub>=0.33. <sup>1</sup>H NMR (500.1 MHz, CDCl<sub>3</sub>) δ, ppm: 7.43 (d, J = 8.5 Hz, 2H), 7.33 (d, J = 8.5 Hz, 2H), 3.86 (s, 3H), 3.41 (s, 2H), 2.23 (s, 6H). <sup>13</sup>C NMR (125.8 MHz, CDCl<sub>3</sub>) δ, ppm: 165.76, 136.63, 129.74, 124.07, 123.94, 63.87, 52.01, 45.33. MS (ESI): Calcd for C<sub>12</sub>H<sub>16</sub>N<sub>3</sub>O<sub>2</sub><sup>+</sup> [M+H]<sup>+</sup> 234.1, found 234.1.

**4-((N,N-Dimethylamino)methyl)bromobenzene (39).** 4-(Bromomethyl)phenylbromide (0.4 g, 1.6



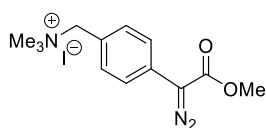
mmol) was suspended in 5 ml of absolute ethanol, dimethylamine in absolute ethanol (5 ml, 5.6 mol/l) was added and the mixture was stirred at room temperature for 30 min. The volatiles were then removed under vacuum and the residue was partitioned between ethyl acetate and 1 M NaOH. The organic phase was washed with 1 M NaOH, brine, dried with anhydrous sodium sulfate and evaporated under vacuum. The remaining semisolid was purified by flash chromatography on Si60 in DCM/methanol to give a yellowish oil (0.24 g, 1.1 mmol, 70 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.30$ .  $^1\text{H NMR}$  (500.1 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 7.44 (d,  $J = 8.3$  Hz, 2H), 7.18 (d,  $J = 8.2$  Hz, 2H), 3.36 (s, 2H), 2.22 (s, 6H).  $^{13}\text{C NMR}$  (125.8 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 138.01, 131.34, 130.72, 120.83, 63.68, 45.34. MS (ESI): Calcd for  $\text{C}_9\text{H}_{12}\text{BrN}^+$   $[\text{M}+\text{H}]^+$  214.0, 216; found. 214.0, 216.0.

**Methyl 2-diazo-2-(4-((dimethylamino)methyl)phenyl)acetate-2- $^{13}\text{C}$  (40).**  $\text{K}_3\text{PO}_4 \cdot \text{H}_2\text{O}$  (0.75 g, 3.0



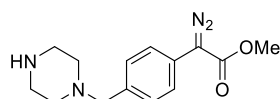
mmol) was mixed with dry toluene (5 ml) in a pressure tube. Compound **3a'** (0.2 g, 0.93 mmol),  $\text{Pd}(\text{OAc})_2$  (0.005 g, 0.002 mmol), 2-di-tert-butylphosphino-2'-methylbiphenyl (0.010 mg, 0.03 mmol) and ethyl acetoacetate-2,4- $^{13}\text{C}_2$  (0.15 g, 1.0 mmol) were then introduced, the tube was flushed with nitrogen and closed. It was then heated up to  $80^\circ\text{C}$  and kept at that temperature for 1 h. The temperature was raised to  $95^\circ\text{C}$  over 1 h, and finally the mixture was kept at  $105^\circ\text{C}$  for 3 h. After cooling down the mixture was centrifuged, the black pellet was washed with 2 x 2 ml of toluene, the organic layers were combined, evaporated and dried under high vacuum to yield a reddish-brown oil. The oil was dissolved in dry methanol (1 ml) and 25 % (w/w) sodium methoxide in methanol (0.5 ml) was added drop-wise. The mixture was left at room temperature for 1 h, then DCM (40 ml) was added and the organic layer was washed with 2 x 30 ml of saturated sodium bicarbonate, dried with anhydrous sodium sulfate, evaporated and dried under high vacuum to give a brown oil. The oil was dissolved in dry acetonitrile (10 ml), p-acetamidobenzensulfonyl azide (0.255 g, 1.1 mmol) and DBU (0.2 g, 1.3 mmol) were then added and the mixture was stirred for 24 h at room temperature. It was then evaporated and purified by flash chromatography on Si60 in DCM/methanol to yield an orange oil (0.083 g, 0.35 mmol, 38 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.33$ .  $^1\text{H NMR}$  (500.1 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 7.41-7.44 (m, 2H), 7.33 (d,  $J = 8.2$  Hz, 2H), 3.86 (s, 3H), 3.41 (s, 2H), 2.23 (s, 6H). MS (ESI): Calcd for  $^{12}\text{C}_{12}^{13}\text{CH}_{16}\text{N}_3\text{O}_2^+$   $[\text{M}+\text{H}]^+$  235.1, found 235.1.

### 1-(4-(1-Diazo-2-methoxy-2-oxoethyl)phenyl)-N,N,N-trimethylmethanaminium iodide (42, Dz3).



Compound **37** (0.024 g, 0.1 mmol) was mixed with neat methyl iodide (0.6 ml, 1.37 g, 9.65 mmol) under nitrogen. The orange oil became light yellow and solidified. The mixture was sonicated, then left for 1 h at room temperature. After that the unreacted methyl iodide was flushed out with nitrogen and the yellow solid was dissolved in acetonitrile-methanol, dried over Si60 and purified in Si60 in DCM/methanol to yield bright yellow crystalline solid (0.026 g, 0.07 mmol, 70 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.25$ .  $^1\text{H NMR}$  (500.1 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 7.65 (d,  $J = 8.6$  Hz, 2H), 7.57 (d,  $J = 8.5$  Hz, 2H), 4.52 (s, 3H), 3.83 (s, 3H), 3.06 (s, 9H).  $^{13}\text{C NMR}$  (125.8 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 165.85, 134.28, 130.07, 125.57, 124.72, 69.33, 69.30, 69.29, 64.32, 53.23, 53.20, 53.16, 52.77. MS (ESI): Calcd for  $\text{C}_{13}\text{H}_{18}\text{N}_3\text{O}_2^+$   $[\text{M}]^+$  248.1, found 248.1.

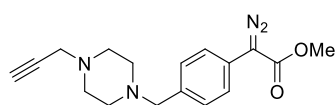
### Methyl 2-diazo-2-(4-(piperazin-1-ylmethyl)phenyl) acetate (44a).



4-(Bromomethyl)- phenylacetic acid (0.4 g, 1.75 mmol) was mixed with dry methanol (3 ml) and dry benzene (3 ml) under nitrogen. Trimethylsilyldiazomethane (2 mol/l in hexanes) was added drop-wise at continuous stirring until stable yellow color of the solution was obtained (the reaction vessel should be vented as gas evolution takes place). The mixture was stirred for 15 min at room temperature, then evaporated and dried under high vacuum to yield the ester as a colorless oil. The oil was dissolved in dry acetonitrile (5 ml), and added to a stirring mixture of *N*-Boc-piperazine (0.335 g, 1.8 mmol) and triethylamine (0.182 mg, 1.8 mmol) in 10 ml of dry acetonitrile under nitrogen. The mixture was stirred until completed as judged by TLC (30 min). The volatiles were removed under vacuum, and the residue was dissolved in dichloromethane (20 ml), washed with  $2 \times 20$  ml of saturated sodium bicarbonate, dried with anhydrous sodium sulfate, evaporated and dried under high vacuum. It was then mixed with *p*-acetamidobenzenesulfonyl azide (0.462 g, 1.93 mmol) in 10 ml of dry acetonitrile under nitrogen. DBU (0.297 g, 1.93 mmol) was added to the solution with continuous stirring and the mixture was stirred for 24 h at room temperature. The mixture was then evaporated under vacuum and the residue purified by flash chromatography on Si60 in DCM/methanol to afford a yellow-orange oil. The oil was dissolved in dry dichloromethane (10 ml), mixed with freshly filtered through a short pad of  $\text{Al}_2\text{O}_3$  2,6-lutidine (0.811 g, 7.6 mmol), and cooled on ice. Trimethylsilyl trifluoromethanesulfonate (0.85 g, 3.8 mmol) was introduced drop-wise and the mixture was stirred until complete as judged by TLC. It was then evaporated to dryness and the residue was purified by column chromatography on Si60 in dichloromethane/methanol to yield the final product as a yellow-brown solid (87 %).  $^1\text{H NMR}$  (500 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ /ppm: 7.50 (d,  $J = 8.5$  Hz, 2H), 7.39 (d,  $J = 8.6$  Hz, 2H), 3.81 (s, 3H), 3.73 (s, 2H), 3.27 – 3.21 (m, 4H), 2.87 – 2.76 (m, 4H).  $^{13}\text{C NMR}$  (126 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ /ppm: 165.41, 132.78,

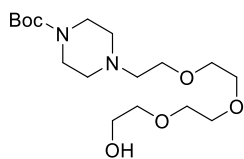
130.11, 125.75, 123.98, 60.95, 51.73, 48.78, 43.57. MS (ESI) calcd for  $C_{14}H_{19}N_4O_2^+$ : (M + H)<sup>+</sup>, 275.22; found: (M + H)<sup>+</sup>, 275.0.

**Methyl 2-diazo-2-(4-((4-(prop-2-yn-1-yl)piperazin-1-yl)methyl)phenyl) acetate (44, Dz4).**



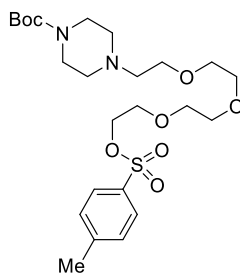
Compound **44a** (80 mg, 0.29 mmol) was mixed with *N,N*-diisopropylethylamine (0.113 mg, 0.86 mmol) and 80 % (w/w) propargyl bromide in toluene (86 mg solution, 0.58 mmol) under nitrogen. The mixture was stirred until complete as judged by TLC (ca. 3 h). The reaction mixture was evaporated and the residue was purified by column chromatography on Si60 in dichloromethane/methanol to yield a red-orange oil (62 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.43$ . IR (film): 2810, 2084, 1702, 1512, 1435, 1150, 1009, 909, 730  $cm^{-1}$ . <sup>1</sup>H NMR (500 MHz,  $CDCl_3$ )  $\delta$ /ppm: 7.42 (d,  $J = 8.5$  Hz, 2H), 7.34 (d,  $J = 8.5$  Hz, 2H), 3.86 (s, 3H), 3.50 (s, 2H), 3.29 (d,  $J = 2.5$  Hz, 2H), 2.55 (d,  $J = 44.8$  Hz, 8H), 2.24 (t,  $J = 2.4$  Hz, 1H). <sup>13</sup>C NMR (126 MHz,  $CDCl_3$ )  $\delta$ /ppm: 165.74, 136.03, 129.79, 124.04, 123.94, 78.94, 73.14, 62.45, 52.93, 52.02, 51.93, 46.83. HRMS (ESI) Calcd for  $C_{17}H_{21}N_4O_2^+$ : (M + H)<sup>+</sup>, 313.1665; found: (M + H)<sup>+</sup>, 313.1659.

**tert-Butyl 4-(2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl)piperazine-1-carboxylate (45b).**



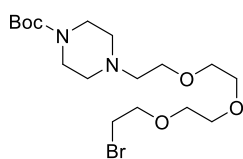
Boc-piperazine (0.362 g, 1.9 mmol) was dissolved in 15 ml of dry acetonitrile under nitrogen.  $K_2CO_3$  (1.0 g, 7.0 mmol) was added, the mixture was heated up to reflux and compound **45a** (prepared as previously described<sup>101</sup>) (0.5 g, 1.9 mmol) as a solution in 5 ml of dry acetonitrile was introduced with continuous stirring. The mixture was refluxed for 18 h, then filtered and the filtrate evaporated under vacuum. The residue was purified by flash chromatography on Si60 in DCM/methanol to afford the target product as a colorless oil (0.551 g, 1.5 mmol, 80 %). TLC (DCM:methanol 20:1 v/v)  $R_f=0.14$ . <sup>1</sup>H NMR (500.1 MHz,  $CDCl_3$ )  $\delta$ , ppm: 3.68 (ddd,  $J = 4.5$  Hz,  $J = 3.5$  Hz,  $J = 2.0$  Hz, 2H), 3.55-3.64 (m, 12H), 3.39-3.41 (m, 4H), 2.55 (dd,  $J = 5.7$  Hz,  $J = 5.7$  Hz, 2H), 2.40-2.42 (m, 4H), 1.41 (s, 9H). <sup>13</sup>C NMR (125.8 MHz,  $CDCl_3$ )  $\delta$ , ppm: 154.69, 79.55, 72.70, 70.58, 70.24, 70.22, 68.64, 61.59, 57.85, 53.32, 43.96, 42.93, 28.40. MS (ESI): Calcd for  $C_{17}H_{34}N_2NaO_6^+$  [M+Na]<sup>+</sup> 385.2, found 385.3.

**tert-Butyl 4-(2-(2-(2-(2-(tosyloxy)ethoxy)ethoxy)ethoxy)ethyl)piperazine-1-carboxylate (45c).**



Compound **45b** (0.550 g, 1.5 mmol) was mixed with triethylamine (0.307 g, 3.0 mmol) and 4-dimethylaminopyridine (0.019 g, 0.16 mmol) in 30 ml of dry acetonitrile under nitrogen. *p*-Toluenesulfonyl chloride (0.318 g, 1.7 mmol) as a solution in 10 ml of dry acetonitrile was introduced at continuous stirring. The mixture was stirred at room temperature for 48 h, then evaporated under vacuum and residue was purified by flash chromatography on Si60 in

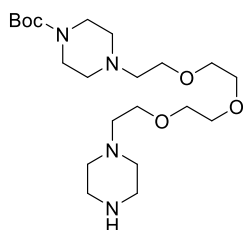
DCM/methanol to yield the target product as a colorless oil (0.517, 1.0 mmol, 66 %). TLC (DCM:methanol 20:1 v/v)  $R_f$ =0.28.  $^1\text{H}$  NMR (500.1 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 7.78 (d,  $J$  = 8.3 Hz, 2H), 7.33 (d,  $J$  = 8.5 Hz, 2H), 4.14 (ddd,  $J$  = 4.5 Hz,  $J$  = 4.0 Hz,  $J$  = 3.0 Hz, 2H), 3.67 (ddd,  $J$  = 5.0 Hz,  $J$  = 4.0 Hz,  $J$  = 3.5 Hz, 2H), 3.57-3.61 (m, 10H), 3.40-3.42 (m, 4H), 2.58 (dd,  $J$  = 6.0 Hz,  $J$  = 6.0 Hz, 2H), 2.41-2.43 (m, 7H), 1.44 (s, 9H).  $^{13}\text{C}$  NMR (125.8 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 154.75, 144.81, 132.99, 129.84, 127.99, 79.58, 70.77, 70.63, 70.52, 70.37, 69.23, 68.86, 68.70, 57.81, 53.37, 44.02, 43.00, 28.43, 21.67. MS (ESI): Calcd for  $\text{C}_{24}\text{H}_{41}\text{N}_2\text{O}_8\text{S}^+$   $[\text{M}+\text{H}]^+$  517.3, found 517.2.



**tert-butyl 4-(2-(2-(2-(2-bromoethoxy)ethoxy)ethoxy)ethyl)piperazine-1-carboxylate (45d).** LiBr (0.695 g, 8.0 mmol) was mixed with dry acetone (12 ml) and heated up to reflux under nitrogen. Compound **45c** (0.517 g, 1.0 mmol) as a solution in 3 ml of dry acetone was added with continuous stirring

and the mixture was refluxed until the reaction was complete as judged by TLC (2 h). The solution was then cooled to room temperature and evaporated under vacuum. The residue was mixed with 30 ml of DCM and the white precipitate that formed was removed by filtration. The filtrate was taken up to dryness and the residue was purified by flash chromatography on Si60 in DCM/methanol to afford the target product as a colorless oil (0.419 g, 1.0 mmol, 98 %). TLC (DCM:methanol 20:1 v/v)  $R_f$ =0.28.  $^1\text{H}$  NMR (500.1 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 3.81 (dd,  $J$  = 6.3 Hz,  $J$  = 6.3 Hz, 2H), 3.61-3.69 (m, 10H), 3.47 (dd,  $J$  = 6.3 Hz,  $J$  = 6.3 Hz, 2H), 3.42-3.44 (m, 4H), 2.60 (dd,  $J$  = 5.9 Hz,  $J$  = 5.9 Hz, 2H), 2.43-2.45 (m, 4H), 1.45 (s, 9H).  $^{13}\text{C}$  NMR (125.8 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 154.77, 79.59, 71.23, 70.70, 70.62, 70.58, 70.42, 68.91, 57.84, 53.38, 40.25, 30.31, 28.44. ESI MS: 425.2, 427.1  $[\text{M}+\text{H}]^+$ , 447.1, 449.1  $[\text{M}+\text{Na}]^+$ . MS (ESI): Calcd for  $\text{C}_{17}\text{H}_{34}\text{BrN}_2\text{O}_5^+$   $[\text{M}+\text{H}]^+$  425.2, 427.2, found 425.2, 427.1; Calcd for  $\text{C}_{17}\text{H}_{33}\text{BrN}_2\text{NaO}_5^+$   $[\text{M}+\text{Na}]^+$  447.2, 449.2, found 447.1, 449.1.

**tert-Butyl 4-(2-(2-(2-(2-(piperazin-1-yl)ethoxy)ethoxy)ethoxy)ethyl)piperazine-1-carboxylate**

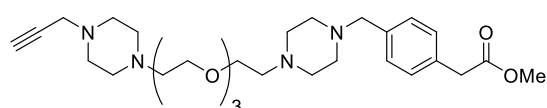


**(45e).** Compound **45d** (0.400 g, 0.93 mmol) was dissolved in 5 ml of dry acetonitrile and added to a stirring mixture of piperazine (0.404 g, 4.7 mmol) and  $\text{K}_2\text{CO}_3$  (0.650 g, 4.7 mmol) in 15 ml of dry acetonitrile. The mixture was refluxed for 18 h under nitrogen. After cooling to room temperature the solids were filtered off, the filtrate was evaporated under vacuum and the residue was purified by flash chromatography on Si60 in DCM/methanol containing 0.5 %

triethylamine. The final product was evaporated two times from toluene and dried under high vacuum to give a colorless oil (0.285 g, 0.66 mmol, 71 %). TLC (DCM:methanol 5:1 v/v containing 0.5 %  $\text{Et}_3\text{N}$ )  $R_f$ =0.22.  $^1\text{H}$  NMR (500.1 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 3.56-3.62 (m, 12H), 3.39-3.41 (m, 4H), 2.87

(dd,  $J = 5.0$  Hz,  $J = 5.0$  Hz, 4H), 2.55 (ddd,  $J = 11.0$  Hz,  $J = 5.5$  Hz,  $J = 5.0$  Hz, 4H), 2.38-2.46 (m, 9H), 1.42 (s, 9H).  $^{13}\text{C}$  NMR (125.8 MHz,  $\text{CDCl}_3$ )  $\delta$ , ppm: 154.72, 79.55, 70.58, 70.39, 70.38, 68.69, 68.82, 58.33, 57.81, 54.71, 53.36, 45.88, 44.00, 43.10. MS (ESI): Calcd for  $\text{C}_{21}\text{H}_{43}\text{N}_4\text{O}_5^+$   $[\text{M}+\text{H}]^+$  431.3, found 431.3; Calcd for  $\text{C}_{21}\text{H}_{42}\text{N}_4\text{NaO}_5^+$   $[\text{M}+\text{Na}]^+$  453.3, found 453.2; Calcd for  $\text{C}_{21}\text{H}_{42}\text{KN}_4\text{O}_5^+$   $[\text{M}+\text{K}]^+$  469.3, found 469.2.

**Methyl 2-(4-(((4-(2-(2-(2-(2-(4-(prop-2-yn-1-yl)piperazin-1-yl)ethoxy)ethoxy)ethoxy)ethyl)piperazin-1-yl)methyl)phenyl)acetate (45f). tert-Butyl 4-(2-(2-(2-(2-(piperazin-1-yl)ethoxy)-**



ethoxy)ethoxy)ethyl)piperazine-1-carboxylate **45e**

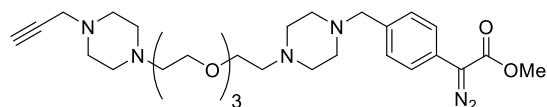
(0.10 g, 0.23 mmol) was mixed with  $\text{K}_2\text{CO}_3$  (0.20 g)

in dry acetonitrile (5 ml) under nitrogen. Methyl 4-

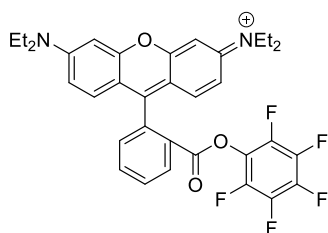
(bromomethyl)phenylacetate **36** (0.056 g, 0.23 mmol) as a solution in 5 ml of acetonitrile was added with continuous stirring. The mixture was stirred at room temperature until the reaction was complete as judged by TLC (2 h). The solids were then filtered off, the filtrate was evaporated to dryness under vacuum and the remaining colorless oil was dissolved in 5 ml of dry DCM. 2,6-Lutidine (0.093 g, 0.87 mmol) was added and the mixture was cooled in an ice-water bath under nitrogen. TMSOTf (0.111 g, 0.5 mmol) was introduced drop-wise with continuous stirring. The ice bath was removed and the mixture was stirred at ambient temperature until complete as judged by TLC (1 h). The solvent was removed under vacuum and the remaining oil was dissolved in dry acetonitrile (5 ml),  $\text{K}_2\text{CO}_3$  (0.26 g) was added and the mixture was stirred for 10 min under nitrogen. Propargyl bromide (0.056 g, 0.36 mmol) as a solution in toluene was then added and the mixture was stirred at room temperature until complete as judged by TLC (1 h). The solvent was taken up to dryness under vacuum and the residue was purified by flash chromatography on Si60 in DCM/methanol to afford the desired product as a colorless oil (0.098 g, 0.18 mmol, 80 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.10$ .  $^1\text{H}$  NMR (500.1 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 7.27 (d,  $J = 8.2$  Hz, 2H), 7.23 (d,  $J = 8.2$  Hz, 2H), 4.53 (br s, 1H), 3.63 (s, 3H), 3.53-3.62 (m, 16H), 3.27 (d,  $J = 2.5$  Hz, 2H), 2.50-2.76 (m, 21 H).  $^{13}\text{C}$  NMR (125.8 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 172.89, 137.38, 134.53, 130.23, 130.08, 79.79, 74.39, 70.78, 70.74, 70.69, 68.15, 67.59, 62.47, 58.08, 57.89, 53.96, 53.89, 52.53, 52.38, 51.63, 46.88, 40.90. MS (ESI): Calcd for  $\text{C}_{29}\text{H}_{47}\text{N}_4\text{O}_5^+$   $[\text{M}+\text{H}]^+$  531.4, found 531.3; Calcd for  $\text{C}_{29}\text{H}_{46}\text{N}_4\text{NaO}_5^+$   $[\text{M}+\text{Na}]^+$  553.3, found 553.3; Calcd for  $\text{C}_{29}\text{H}_{46}\text{KN}_4\text{O}_5^+$   $[\text{M}+\text{K}]^+$  569.3, found 569.3.



**Methyl 2-diazo-2-(4-((4-(2-(2-(2-(2-(4-(prop-2-yn-1-yl)piperazin-1-yl)ethoxy)ethoxy)ethoxy)ethyl)piperazin-1-yl)methyl)phenyl)acetate (45, Dz5).**



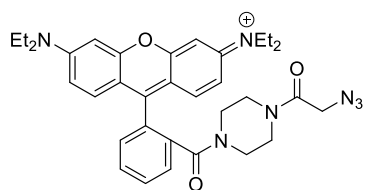
Compound **45f** (0.090 g, 0.17 mmol) was mixed with p-acetamidobenzensulfonyl azide (0.062 g, 0.26 mmol) in 5 ml of dry acetonitrile under nitrogen. DBU (0.103 g, 0.68 mmol) was introduced drop-wise with continuous stirring. The mixture was stirred for 48 h at room temperature, then evaporated to dryness under vacuum. Purification of the residue by flash chromatography on Si60 in DCM/methanol yielded the target product as an orange-yellow oil (0.028 g, 0.05 mmol, 30 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.22$ .  $^1\text{H NMR}$  (500.1 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 7.44 (d,  $J = 8.5$  Hz, 2H), 7.33 (d,  $J = 8.5$  Hz, 2H), 3.81 (s, 3H), 3.49-3.54 (m, 12H), 3.44 (s, 2H), 3.22 (d,  $J = 2.5$  Hz, 2H), 2.39-2.49 (m, 21H).  $^{13}\text{C NMR}$  (125.8 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 166.48, 137.53, 130.49, 125.19, 124.83, 80.19, 74.15, 71.11, 71.01, 69.53, 69.47, 62.90, 58.57, 58.51, 54.42, 54.24, 53.92, 52.61, 52.52, 47.12. MS (ESI): Calcd for  $\text{C}_{29}\text{H}_{45}\text{N}_6\text{O}_5^+$   $[\text{M}+\text{H}]^+$  557.3, found 557.4; Calcd for  $\text{C}_{29}\text{H}_{44}\text{N}_6\text{NaO}_5^+$   $[\text{M}+\text{Na}]^+$  579.3, found 579.3; Calcd for  $\text{C}_{29}\text{H}_{44}\text{KN}_6\text{O}_5^+$   $[\text{M}+\text{K}]^+$  595.3, found 595.3.



**Rhodamine B 2,3,4,5,6-pentafluorophenyl ester (46a).**

Rhodamine B (0.600 g, 1.35 mmol), 2,3,4,5,6-pentafluorophenol (0.250 g, 1.36 mmol) and 4-(N,N-dimethylamino)pyridine (0.021 g, 0.17 mmol) were dissolved in 20 ml of dry DCM under nitrogen and N,N'-dicyclohexylcarbodiimide (0.327 g, 1.59 mmol) as a solution in 2 ml of DCM was added with continuous stirring. The mixture was stirred at room temperature until complete as judged by TLC (1.5 h). It was then evaporated and the residue purified by column chromatography on Si60 in DCM/methanol to afford the target product as a deep-purple solid (0.610 g, 1.0 mmol, 74 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.20$ .  $^1\text{H NMR}$  (500.1 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 8.53 (dd,  $J = 8.0$  Hz,  $J = 1.0$  Hz, 1H), 8.02 (ddd,  $J = 7.6$  Hz,  $J = 7.6$  Hz,  $J = 1.3$  Hz, 1H), 7.93 (ddd,  $J = 7.8$  Hz,  $J = 7.8$  Hz,  $J = 1.3$  Hz, 1H), 7.53 (dd,  $J = 7.6$  Hz,  $J = 1.0$  Hz, 1H), 7.12 (d,  $J = 9.5$  Hz, 2H), 6.98 (dd,  $J = 9.6$  Hz,  $J = 2.5$  Hz, 2H), 6.82 (d,  $J = 2.5$  Hz, 2H), 3.62 (q,  $J = 7.2$  Hz, 8H), 1.23 (t,  $J = 7.1$  Hz, 12H).  $^{13}\text{C NMR}$  (125.8 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: 162.47, 158.67, 157.44, 156.63, 142.91 (m, CF), 140.91 (m, CF), 139.88 (m, CF), 137.93 (m, CF), 135.96, 135.79, 133.05, 131.90, 131.82, 131.77, 127.45, 115.47, 114.14, 96.95, 46.66, 12.72.  $^{19}\text{F NMR}$  (376.5 MHz,  $\text{CD}_3\text{CN}$ )  $\delta$ , ppm: -154.76 (d,  $J = 16.8$  Hz), -159.47 (t,  $J = 21.0$  Hz), -164.16 (dd,  $J = 16.9$  Hz,  $J = 21.9$  Hz). MS (ESI): Calcd for  $\text{C}_{34}\text{H}_{30}\text{F}_5\text{N}_2\text{O}_3^+$   $[\text{M}]^+$  609.2, found 609.3.

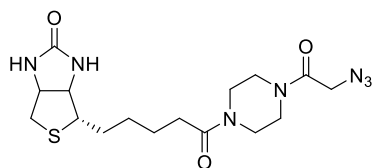
**1-(2-Azidoacetyl)-4-rhodamine B piperazine (46).** Compound **46a** (0.035 g, 0.06 mmol) was



dissolved in 5 ml of dry acetonitrile and N-(2-azidoacetyl)piperazine<sup>102</sup> (0.014 g, 0.08 mmol) as a solution in 1 ml of acetonitrile was added with continuous stirring under nitrogen. The mixture was stirred for 24 h at room temperature, then evaporated under vacuum and the residue purified by column

chromatography on Si60 in DCM/methanol to afford a deep purplish-violet waxy solid (0.016 g, 0.03 mmol, 48 %). TLC (DCM:methanol 5:1 v/v)  $R_f=0.32$ . <sup>1</sup>H NMR (500.1 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ , ppm: 7.71-7.78 (m, 3H), 7.52-7.53 (m, 1H), 7.10-7.16 (m, 4H), 6.95 (d,  $J = 2.3$  Hz), 4.11-4.13 (m, 3H), 3.66 (q,  $J = 7.2$  Hz, 8H), 3.39-3.41 (m, 3H), 3.22-3.24 (m, 1H), 2.65-2.69 (m, 3H), 1.20 (t,  $J = 7.2$  Hz, 12H). <sup>13</sup>C NMR (125.8 MHz, DMSO-*d*<sub>6</sub>)  $\delta$ , ppm: 166.07, 165.77, 157.05, 155.54, 155.11, 135.13, 131.77, 130.82, 130.74, 130.71, 130.45, 129.82, 127.54, 114.30, 113.02, 95.90, 49.67, 45.40, 45.36, 45.15, 45.08, 42.31, 12.47. MS (ESI): Calcd for C<sub>34</sub>H<sub>40</sub>N<sub>7</sub>O<sub>3</sub><sup>+</sup> [M]<sup>+</sup> 594.3, found 594.4.

**1-(2-Azidoacetyl)-4-(D-biotinyl)piperazine (47).** D-Biotin 4-nitrophenyl ester (0.068 g, 0.19 mmol)



was suspended in 5 ml of dry acetonitrile and N-(2-azidoacetyl)piperazine<sup>102</sup> (0.042 g, 0.25 mmol) as a solution in 1 ml of acetonitrile was added with continuous stirring under nitrogen.

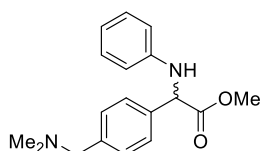
The mixture was stirred for 24 h at room temperature, then evaporated under vacuum and the residue purified by column chromatography on Si60 in DCM/methanol to afford a colorless oil which was crystallized from chloroform (0.062 g, 0.16 mmol, 84 %). TLC (DCM:methanol 10:1 v/v)  $R_f=0.23$ . <sup>1</sup>H NMR (500.1 MHz, CD<sub>3</sub>CN)  $\delta$ , ppm: 5.42 (br s, 1H), 5.15 (br s, 1H), 4.41 (dddd,  $J = 6.5$  Hz,  $J = 5.0$  Hz,  $J = 2.5$  Hz,  $J = 1.0$  Hz, 1H), 4.24 (ddd,  $J = 8.0$  Hz,  $J = 4.5$  Hz,  $J = 2.0$  Hz, 1H), 3.99 (s, 2H), 3.47-3.56 (m, 6H), 3.28-3.34 (m, 2H), 3.17 (ddd,  $J = 7.9$  Hz,  $J = 6.9$  Hz,  $J = 4.5$  Hz, 1H), 2.88 (dd,  $J = 12.7$  Hz,  $J = 5.0$  Hz, 1H), 2.63 (d,  $J = 12.7$  Hz, 1H), 2.35 (dd,  $J = 8.1$  Hz,  $J = 7.0$  Hz, 2H), 1.65-1.72 (m, 1H), 1.53-1.63 (m, 3H), 1.38-1.44 (m, 2H). <sup>13</sup>C NMR (125.8 MHz, CD<sub>3</sub>CN)  $\delta$ , ppm: 172.37, 167.13, 163.83, 62.30, 60.71, 56.30, 50.99, 45.75, 45.40, 45.11, 42.63, 42.32, 41.74, 41.66, 41.12, 33.19, 29.10, 29.04, 25.75. ESI MS: 418.2 [M+Na]<sup>+</sup>. ESI MS: Calcd for C<sub>16</sub>H<sub>25</sub>N<sub>7</sub>NaO<sub>3</sub>S<sup>+</sup> [M+Na]<sup>+</sup> 418.2, found 418.2.

### 5.3.1. Monoalkylation of small anilines with copper(I) carbenoids derived from Dz1 in aqueous media

Compound **37** (**Dz1**, 0.082 mmol) and the respective aniline or aminopyridine (0.090 mmol) were mixed in *tert*-butanol (0.33 ml) under nitrogen, 500 mM MES buffer, pH 6.0 (0.65 ml), 50 mM CuSO<sub>4</sub> in water (0.016 ml) and water (0.47 ml) were then added, the mixture was stirred, and 200 mM sodium

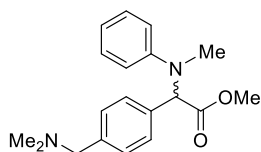
ascorbate in water (0.16 ml) was introduced. The mixture was stirred for 1 h (compounds 3a-g) or overnight (compounds 3h and 3i) at room temperature. Equal volume of 10 % (w/v)  $K_2CO_3$  was added, and the mixture was extracted with  $3 \times 4$  ml of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated to dryness under vacuum and the residue purified by column chromatography on Si60 in DCM/methanol to yield the desired product.

**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-(phenylamino) acetate (53a).** Light yellow oil (85



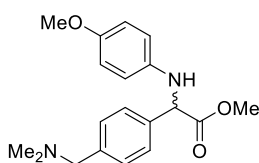
%). TLC (DCM-methanol 20:1 v/v)  $R_f=0.38$ . IR (film) 2768, 1735, 1601, 1504, 1432, 1309, 1253, 1170, 1018, 857, 746, 691, 508  $cm^{-1}$ .  $^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta/ppm$ : 7.45 (d,  $J = 8.1$  Hz, 2H), 7.31 (d,  $J = 8.1$  Hz, 2H), 7.12 (dd,  $J = 8.5, 7.4$  Hz, 2H), 6.70 (dd,  $J = 7.3, 7.3$  Hz, 1H), 6.56 (d,  $J = 7.7$  Hz, 2H), 5.07 (d,  $J = 6.0$  Hz, 1H), 4.93 (d,  $J = 5.8$  Hz, 1H), 3.73 (s, 3H), 3.44 (s, 2H), 2.26 (s, 6H).  $^{13}C$  NMR (126 MHz,  $CDCl_3$ )  $\delta/ppm$ : 172.49, 146.07, 138.71, 136.65, 129.80, 129.36, 127.35, 118.23, 113.50, 63.93, 60.62, 52.93, 45.39. HRMS (ESI) Calcd. for  $C_{18}H_{23}N_2O^+$ : (M + H) $^+$ , 299.1760; found: (M + H) $^+$ , 299.1754.

**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-(methyl(phenyl)amino) acetate (53b).**



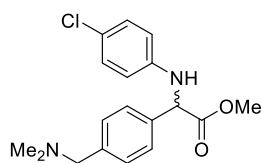
Colorless oil (82 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.33$ . IR (film) 2767, 1742, 1570, 1503, 1169, 1106, 1002, 749, 691  $cm^{-1}$ .  $^1H$  NMR (400 MHz,  $CDCl_3$ )  $\delta/ppm$ : 7.30 (d,  $J = 8.2$  Hz, 2H), 7.27 – 7.18 (m, 4H), 6.84 (d,  $J = 8.0$  Hz, 2H), 6.80 – 6.74 (m, 1H), 5.62 (s, 1H), 3.75 (s, 3H), 3.44 (s, 2H), 2.75 (s, 3H), 2.25 (s, 6H).  $^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta/ppm$ : 172.41, 149.85, 138.34, 134.78, 129.47, 129.31, 128.43, 118.08, 113.45, 65.48, 63.78, 52.02, 45.26, 34.46. HRMS (ESI) Calcd. for  $C_{19}H_{25}N_2O_2^+$ : (M + H) $^+$ , 313.1916; found: (M + H) $^+$ , 313.1914.

**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-((4-methoxyphenyl)amino) acetate (53c).**



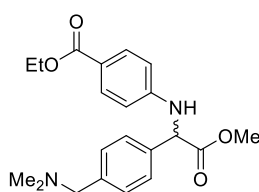
Colorless oil (75 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.35$ . IR (film) 1735, 1510, 1235, 1033, 818  $cm^{-1}$ .  $^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta/ppm$ : 7.44 (d,  $J = 8.1$  Hz, 2H), 7.31 (d,  $J = 8.1$  Hz, 2H), 6.72 (d,  $J = 8.9$  Hz, 2H), 6.53 (d,  $J = 8.9$  Hz, 2H), 5.01 (d,  $J = 6.4$  Hz, 1H), 4.64 (d,  $J = 6.4$  Hz, 1H), 3.72 (s, 3H), 3.70 (s, 3H), 3.44 (s, 2H), 2.26 (s, 6H).  $^{13}C$  NMR (126 MHz,  $CDCl_3$ )  $\delta/ppm$ : 172.61, 152.52, 140.21, 138.41, 136.75, 129.68, 127.27, 114.84, 114.77, 63.80, 61.43, 55.69, 52.74, 45.25. HRMS (ESI) Calcd. for  $C_{19}H_{25}N_2O_3^+$ : (M + H) $^+$ , 329.1865; found: (M + H) $^+$ , 329.1865.

**Methyl 2-((4-chlorophenyl)amino)-2-(4-((dimethylamino)methyl)phenyl) acetate (53d).**



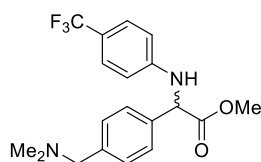
Colorless oil (77 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.35$ . IR (film) 1735, 1598, 1498, 1312, 1171, 1018, 815  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 7.42 (d,  $J = 8.1$  Hz, 2H), 7.31 (d,  $J = 8.2$  Hz, 2H), 7.05 (d,  $J = 8.9$  Hz, 2H), 6.46 (d,  $J = 8.9$  Hz, 2H), 5.02 (d,  $J = 6.0$  Hz, 1H), 4.97 (d,  $J = 5.9$  Hz, 1H), 3.73 (s, 3H), 3.44 (s, 2H), 2.26 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 172.08, 144.44, 138.73, 136.06, 129.75, 129.09, 127.19, 122.75, 114.50, 63.78, 60.46, 52.92, 45.28. HRMS (ESI) Calcd. for  $\text{C}_{18}\text{H}_{22}\text{ClN}_2\text{O}_2^+$ : (M + H) $^+$ , 333.1370; found: (M + H) $^+$ , 333.1363.

**Ethyl 4-((1-(4-((dimethylamino)methyl)phenyl)-2-methoxy-2-oxoethyl)amino) benzoate (53e).**



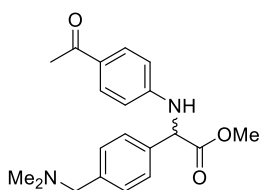
Colorless oil (70 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.33$ . IR (film) 1737, 1697, 1603, 1523, 1265, 1169, 1098, 1018, 839, 769, 698  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 7.81 (d,  $J = 8.8$  Hz, 2H), 7.42 (d,  $J = 8.1$  Hz, 2H), 7.31 (d,  $J = 8.2$  Hz, 2H), 6.51 (d,  $J = 8.9$  Hz, 2H), 5.41 (d,  $J = 5.8$  Hz, 1H), 5.12 (d,  $J = 5.8$  Hz, 1H), 4.28 (q,  $J = 7.1$  Hz, 2H), 3.74 (s, 3H), 3.43 (s, 2H), 2.25 (s, 6H), 1.32 (t,  $J = 7.1$  Hz, 3H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 171.74, 166.67, 149.43, 138.90, 135.70, 131.41, 129.79, 127.15, 119.70, 112.32, 77.30, 77.05, 76.79, 63.75, 60.25, 59.81, 53.04, 45.28, 14.42. HRMS (ESI) Calcd. for  $\text{C}_{21}\text{H}_{27}\text{N}_2\text{O}_4^+$ : (M + H) $^+$ , 371.1971; found: (M + H) $^+$ , 371.1971.

**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-((4-(trifluoromethyl)phenyl)amino)acetate (53f).**



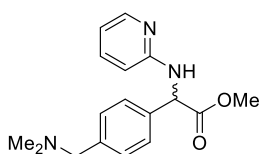
Colorless oil (74 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.30$ . IR (film) 1737, 1616, 1532, 1315, 1105, 1064, 825  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 7.42 (d,  $J = 8.1$  Hz, 2H), 7.35-7.32 (m, 4H), 6.55 (d,  $J = 8.5$  Hz, 2H), 5.33 (d,  $J = 5.7$  Hz, 1H), 5.08 (d,  $J = 5.7$  Hz, 1H), 3.75 (s, 3H), 3.44 (s, 2H), 2.26 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 171.74, 148.21, 138.83, 135.65, 129.79, 127.11, 126.57 (q,  $J=4$  Hz), 124.76 (q,  $J=271$  Hz), 119.61 (q,  $J=33$  Hz), 112.58, 63.71, 59.88, 53.02, 45.24. HRMS (ESI) Calcd. for  $\text{C}_{19}\text{H}_{22}\text{F}_3\text{N}_2\text{O}_2^+$ : (M + H) $^+$ , 367.1633; found: (M + H) $^+$ , 367.1633.

**Methyl 2-((4-acetylphenyl)amino)-2-(4-((dimethylamino)methyl)phenyl) acetate (53g).**



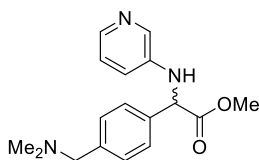
Colorless oil (75 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.30$ . IR (film) 2769, 1737, 1660, 1594, 1524, 1420, 1357, 1270, 1171, 1018, 825, 591  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 7.75 (d,  $J = 8.8$  Hz, 2H), 7.41 (d,  $J = 8.1$  Hz, 2H), 7.32 (d,  $J = 8.2$  Hz, 2H), 6.52 (d,  $J = 8.9$  Hz, 2H), 5.51 (d,  $J = 5.7$  Hz, 1H), 5.13 (d,  $J = 5.8$  Hz, 1H), 3.75 (s, 3H), 3.43 (s, 2H), 2.45 (s, 3H), 2.25 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta/\text{ppm}$ : 196.38, 171.63, 149.68, 139.01, 135.57, 130.65, 129.83, 127.52, 127.13, 112.32, 63.75, 59.72, 53.10, 45.28, 26.05. HRMS (ESI) Calcd. for  $\text{C}_{20}\text{H}_{25}\text{N}_2\text{O}_3^+$ : (M + H) $^+$ , 341.1865; found: (M + H) $^+$ , 341.1863.

**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-(pyridin-2-ylamino) acetate (53h).**

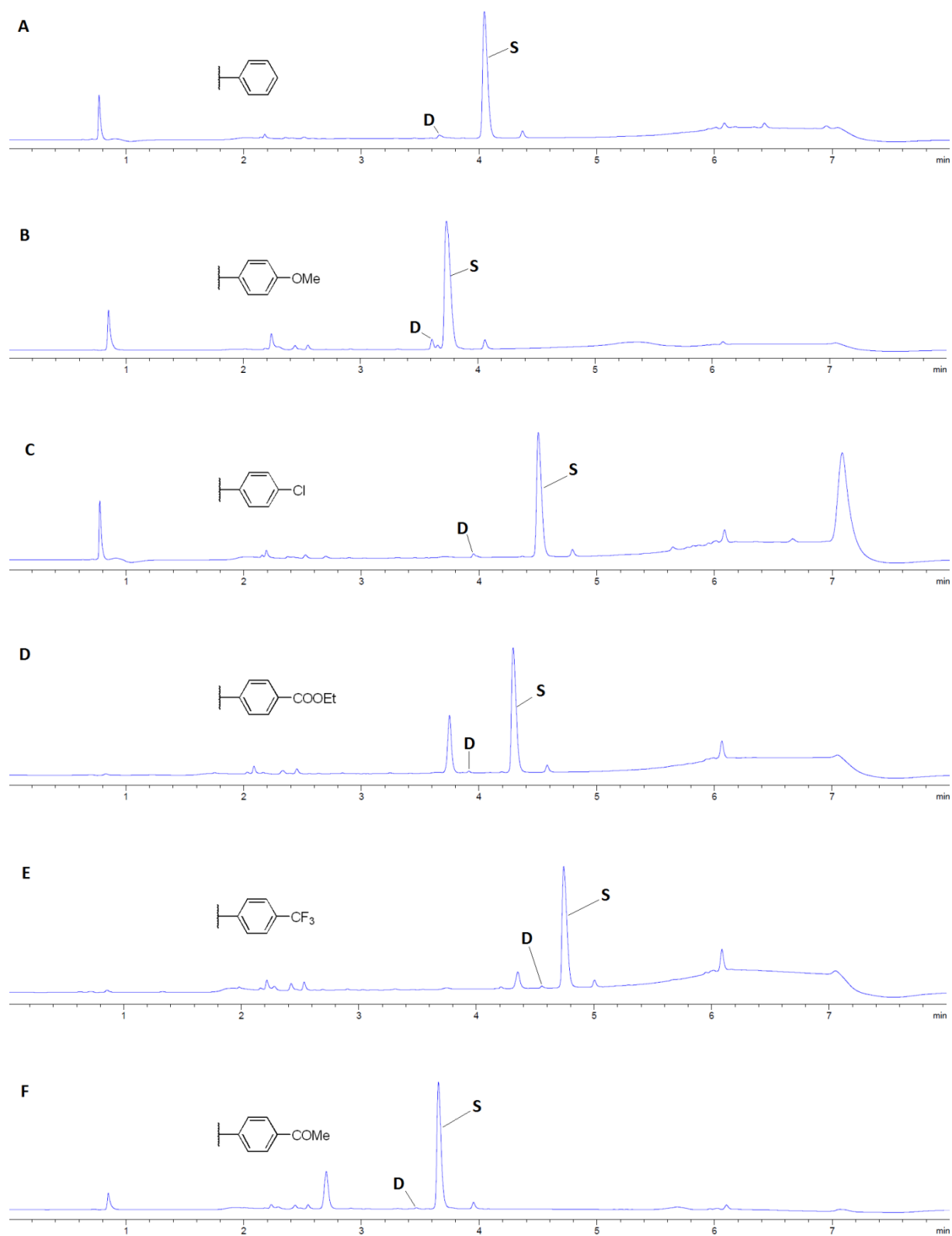


Colorless oil (51 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.25$ . IR (film) 2768, 1737, 1600, 1480, 1169, 1018, 772  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 7.98 (dd,  $J = 5.6, 1.9$  Hz, 1H), 7.41 (d,  $J = 8.0$  Hz, 3H), 7.32 (d,  $J = 8.3$  Hz, 2H), 6.71 – 6.53 (m, 2H), 5.88 (s, 1H), 5.55 (d,  $J = 7.0$  Hz, 1H), 3.64 (s, 3H), 3.41 (s, 2H), 2.17 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 172.70, 157.46, 147.45, 139.46, 137.11, 136.27, 129.35, 127.57, 113.45, 109.04, 63.20, 58.26, 51.91, 44.48. HRMS (ESI) Calcd. for  $\text{C}_{17}\text{H}_{22}\text{N}_3\text{O}_2^+$ : (M + H) $^+$ , 300.1712; found: (M + H) $^+$ , 300.1706.

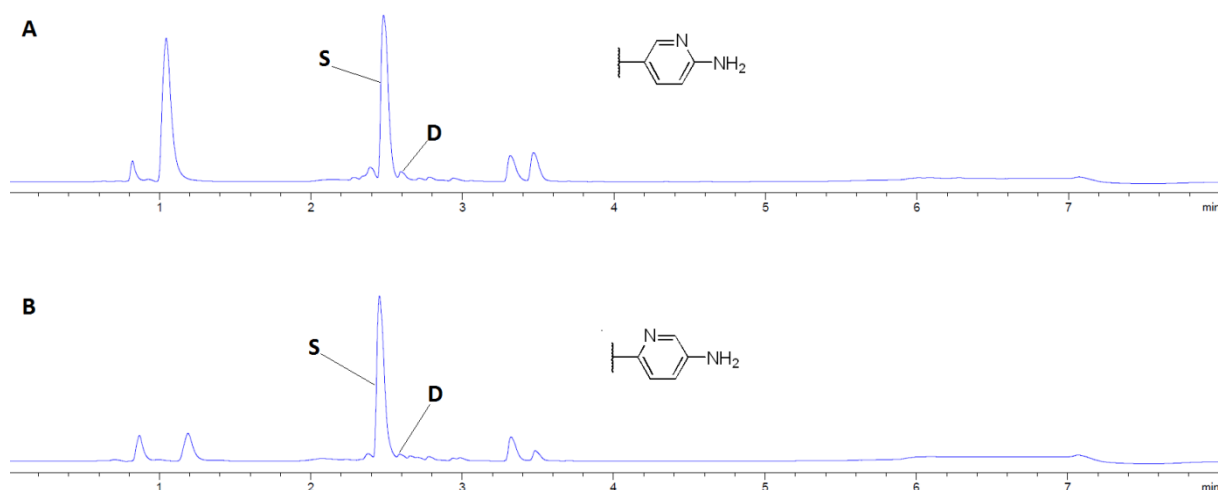
**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-(pyridin-3-ylamino) acetate (53i).**



oil (76 %). TLC (DCM-methanol 20:1 v/v)  $R_f=0.023$ . IR (film) 2769, 1737, 1587, 1480, 1171, 1017, 795, 708  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 8.06 (s, 1H), 7.89 (s, 1H), 7.44 (d,  $J = 8.1$  Hz, 2H), 7.34 (d,  $J = 8.3$  Hz, 2H), 7.05 (s, 1H), 6.92 (d,  $J = 6.6$  Hz, 1H), 5.48 (d,  $J = 7.1$  Hz, 1H), 5.20 (d,  $J = 7.3$  Hz, 1H), 3.67 (s, 3H), 3.44 (s, 2H), 2.19 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 171.86, 139.08, 138.83, 136.26, 136.22, 129.57, 127.36, 123.71, 119.20, 62.95, 59.48, 52.36, 44.31. HRMS (ESI) Calcd. for  $\text{C}_{17}\text{H}_{22}\text{N}_3\text{O}_2^+$ : (M + H) $^+$ , 300.1712; found: (M + H) $^+$ , 300.1709.



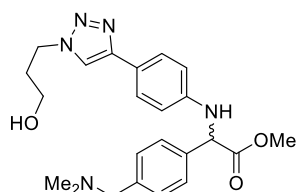
**Figure E1** | UPLC trace of the Cu(I)-catalyzed alkylation reaction of a series of *p*-substituted anilines with Dz1 (Sections A-F correspond to entries 1 and 3-7 in Table 7). The singly (S) and doubly alkylated (D) anilines were identified by ESI-MS



**Figure E2** | UPLC trace of the Cu(I)-catalyzed alkylation reaction of a series of 2- and 3-aminopyridine with Dz1 (Sections A and B correspond to entries 8 and 9 in Table 7). The singly (S) and doubly alkylated (D) anilines were identified by ESI-MS.

### 5.3.2. Auto-catalytic tandem CuAAC/N-H insertion with small anilines

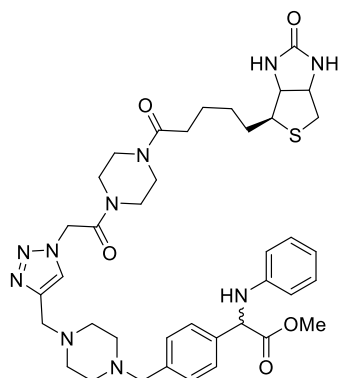
**Methyl 2-(4-((dimethylamino)methyl)phenyl)-2-((4-(1-(3-hydroxypropyl)-1H-1,2,3-triazol-4-yl)phenyl)amino) acetate (55).** Compound **37** (**Dz1**, 0.064 mmol), 4-ethynylaniline (0.071 mmol) and



3-azidopropan-1-ol<sup>86</sup> (0.077 mmol) were mixed in *tert*-butanol (0.45 ml) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.514 ml), 50 mM CuSO<sub>4</sub> in water (0.013 ml) and water (0.245 ml) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.064 ml)

was introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K<sub>2</sub>CO<sub>3</sub> was added, and the mixture was extracted with 3× 3 ml of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue purified by column chromatography on Si60 in dichloromethane/methanol to yield the desired product as a colorless oil (62 %). TLC (DCM-methanol 20:1 v/v) R<sub>f</sub>=0.10. IR (film) 2948, 1735, 1617, 1564, 1502, 1456, 1317, 1172, 1041, 798 cm<sup>-1</sup>. <sup>1</sup>H NMR (500 MHz, MeOD) δ/ppm: 8.09 (s, 1H), 7.59 – 7.46 (m, 4H), 7.34 (d, *J* = 8.2 Hz, 2H), 6.70 (d, *J* = 8.8 Hz, 2H), 5.23 (s, 1H), 4.50 (t, *J* = 7.0 Hz, 2H), 3.71 (s, 3H), 3.58 (t, *J* = 6.1 Hz, 2H), 3.55 (s, 2H), 2.29 (s, 6H), 2.17 – 2.08 (m, 2H). <sup>13</sup>C NMR (126 MHz, MeOD) δ/ppm: 173.80, 149.32, 148.29, 138.70, 138.13, 131.23, 128.77, 127.63, 121.07, 120.92, 114.75, 64.22, 61.50, 59.31, 53.05, 48.26, 45.00, 33.99. HRMS (ESI) Calcd. for C<sub>23</sub>H<sub>30</sub>N<sub>5</sub>O<sub>3</sub><sup>+</sup>: (M + H)<sup>+</sup>, 424.2349; found: (M + H)<sup>+</sup>, 424.2343.

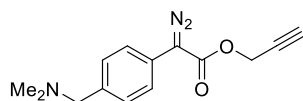
**Methyl 2-(4-((4-((1-(2-oxo-2-(4-(5-((4S)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoyl)piperazin-1-yl)ethyl)-1H-1,2,3-triazol-4-yl)methyl)piperazin-1-yl)methyl)phenyl)-2-(phenyl-amino) acetate (56).** Compound **44 (Dz4)**, 0.058 mmol) and aniline (0.063 mmol) were mixed in *tert*-



butanol (0.406 ml) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.464 ml), 50 mM CuSO<sub>4</sub> in water (0.012 ml) and 1-(2-azidoacetyl)-4-(D-biotinyl)piperazine **47** (0.069 mmol) dissolved in water (0.220 ml) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.058 ml) was introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K<sub>2</sub>CO<sub>3</sub> was added, and the mixture was extracted with 3 × 3 ml of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue

purified by column chromatography on Si60 in dichloromethane/methanol to yield the desired product as a colorless oil (70 %). TLC (DCM-methanol 4:1 v/v) R<sub>f</sub>=0.25. IR (film) 1665, 1500, 1421, 1215, 1006, 825, 751, 693 cm<sup>-1</sup>. <sup>1</sup>H NMR (500 MHz, MeOD) δ/ppm: 7.89 (s, 1H), 7.48 (d, *J* = 8.1 Hz, 2H), 7.33 (d, *J* = 8.2 Hz, 2H), 7.07 (dd, *J* = 8.5, 7.5 Hz, 2H), 6.66-6.62 (m, 3H), 5.51 (s, 2H), 5.16 (s, 1H), 4.56 – 4.45 (m, 1H), 4.32 (dd, *J* = 7.7, 4.5 Hz, 1H), 3.77 – 3.58 (m, 13H), 3.54 (s, 2H), 3.22 (dd, *J* = 7.9, 5.5 Hz, 1H), 2.94 (dd, *J* = 12.8, 4.9 Hz, 1H), 2.79 – 2.41 (m, 10H), 1.84 – 1.58 (m, 4H), 1.55 – 1.45 (m, 2H). <sup>13</sup>C NMR (126 MHz, MeOD) δ/ppm: 172.73, 165.13, 165.04, 164.70, 146.65, 142.85, 137.03, 136.96, 129.66, 128.61, 127.19, 125.95, 117.48, 113.27, 61.95, 61.88, 60.32, 60.23, 55.63, 52.11, 51.90, 51.58, 50.61, 48.45, 45.00, 44.79, 44.62, 44.27, 41.99, 41.60, 41.10, 40.85, 39.66, 32.23, 32.18, 28.48, 28.15, 24.84. HRMS (ESI) Calcd. for C<sub>39</sub>H<sub>53</sub>N<sub>10</sub>O<sub>5</sub>S<sup>+</sup>: (M + H)<sup>+</sup>, 773.3921; found: (M + H)<sup>+</sup>, 773.3921.

**Prop-2-yn-1-yl 2-diazo-2-(4-((dimethylamino)methyl)phenyl) acetate (57, Dz8).** 4-(Bromomethyl)-



phenylacetic acid **35** (0.117 g, 0.51 mmol) was mixed with propargyl alcohol (0.084 g, 1.5 mmol) and 1,2-dichloroethane (0.5 ml) were mixed in a 4-ml screw top vial with a teflon cap. A drop of concentrated sulfuric acid was added and the vial was closed tightly and heated at

80°C for 1h. The mixture was then diluted with 3 ml of dichloromethane, washed with 2 × 4 ml of saturated sodium bicarbonate, dried with anhydrous sodium sulfate and evaporated to dryness under vacuum. The residue was passed through a small pad of Si60 in dichloromethane/methanol 100:1 and the first fraction was evaporated and dried under high vacuum. The residue was mixed with ethanol (2 ml) and 4 ml of 5.6 M dimethylamine in ethanol were added. The mixture was kept for 1 h at room temperature, then taken up to dryness, re-dissolved in 5 ml of dichloromethane, washed with 2 × 5 ml of saturated sodium carbonate solution, then dried with anhydrous sodium sulfate, taken up to dryness



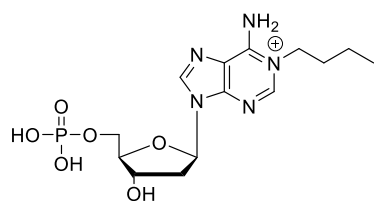
and dried under high vacuum. The residue was dissolved in dry acetonitrile (5 ml) under nitrogen, and *p*-acetamidobenzenesulfonyl azide (0.135 g, 0.56 mmol) and DBU (0.154 g, 1.0 mmol) were then added, and the mixture was stirred for 24 h at room temperature. The mixture was then evaporated under vacuum and the residue purified by flash chromatography on Si60 in dichloromethane/methanol to afford a yellow-orange oil (55 %). TLC (DCM-methanol 4:1 v/v)  $R_f=0.25$ . IR (film) 2768, 2084, 1701, 1513, 1336, 1240, 1144, 1036, 808  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 7.44 (d,  $J = 8.5$  Hz, 2H), 7.34 (d,  $J = 8.7$  Hz, 2H), 4.86 (d,  $J = 2.5$  Hz, 2H), 3.38 (s, 2H), 2.82 (t,  $J = 2.5$  Hz, 1H), 2.16 (s, 6H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CD}_3\text{CN}$ )  $\delta/\text{ppm}$ : 164.33, 137.60, 129.49, 123.96, 123.75, 77.93, 75.51, 63.19, 52.12, 44.53. HRMS (ESI) Calcd. for  $\text{C}_{17}\text{H}_{22}\text{N}_3\text{O}_2^+$ : (M + H) $^+$ , 258.1243; found: (M + H) $^+$ , 258.1237.

### 5.3.3. Direct alkylation of 2'-deoxynucleoside-5'-monophosphates

Typically 2'-deoxyadenosine-5'-monophosphate or 2'-deoxycytidine-5'-monophosphate sodium salt (0.1-0.2 mmol) was suspended in DMSO (0.5 ml) under argon and the respective alkyl bromide or iodide (2-5 eq) was added drop-wise at continuous stirring. The mixture was stirred at room temperature. The reaction progress was checked by diluting small aliquots with 100 mM triethylammonium acetate buffer, pH 7.1/acetonitrile 1:1 and injecting into ESI MS. The mixture was then precipitated with 40-80 volumes of acetone at room temperature, the precipitate was recovered by centrifugation and washed twice by resuspension in acetone and centrifugation. It was then dried under high vacuum and purified by preparative RP-HPLC in water/acetonitrile. The appropriate fractions were combined and freeze-dried to afford the desired alkylated products as white solids.

The alkylation of the nucleobase was confirmed by ESI tandem MS of the target product, where the mass for the alkylated nucleobase was found in all cases. For compound **61c** only the cation of the singly-alkylated base was observed. This together with the splitting pattern of the propargyl  $\text{CH}_2$ -protons due to spin-spin coupling to the 5'-phosphorus suggests the second alkylation took place on the phosphate.

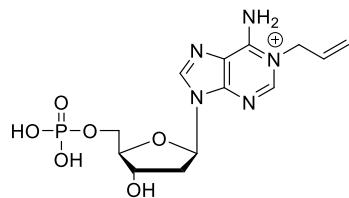
**$\text{N}^1$ -(*n*-Butyl)-2'-deoxyadenosine-5'-monophosphate (60c)**. The reaction was done with 1.1 mmol nucleoside derivative and 5 eq 1-iodobutane, reaction time 3 days.



The purification was carried out by semi-preparative HPLC in 100 mM triethylammonium acetate buffer, pH 7.1/acetonitrile. The final product was re-dissolved in water and freeze-dried three times to remove the triethylammonium acetate, however, some residual amount was still present. Yield 2 mg (0.005 mmol, 4.5 %).  $^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta/\text{ppm}$ : 8.63 (s, 1H), 8.52 (s, 1H), 6.55 (t,  $J = 6.7$  Hz, 1H), 4.77 – 4.71 (m, 1H), 4.32 (t,  $J = 7.5$  Hz, 2H), 4.29-4.27 (m,

1H), 4.02 (dd,  $J = 5.0, 3.6$  Hz, 2H), 2.87 (ddd,  $J = 14.0, 6.9, 6.1$  Hz, 1H), 2.61 (ddd,  $J = 14.1, 6.5, 3.7$  Hz, 1H), 1.88-1.82 (m, 2H), 1.43-1.36 (m, 2H), 0.93 (t,  $J = 7.4$  Hz, 3H). MS (ESI) Calcd. for  $C_{14}H_{23}N_5O_6P^+$ :  $[M]^+$ , 388.1; found: 388.3.

**N<sup>1</sup>-Allyl-2'-deoxyadenosine-5'-monophosphate (60d).** The reaction was done with 2 mmol

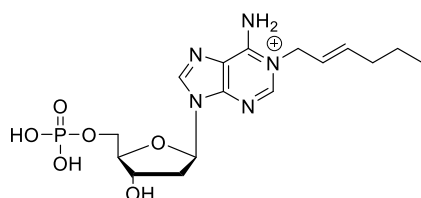


nucleoside derivative and 4 eq allyl bromide, reaction time 20 h.

Yield 7 mg (0.02 mmol, 9 %). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 8.63 (s, 1H), 8.55 (s, 1H), 6.59 (t,  $J = 6.7$  Hz, 1H), 6.12-6.06 (m, 1H), 5.46-5.43 (m, 1H), 5.21-5.18 (m, 1H), 5.00-4.99 (m, 2H), 4.78-4.76 (m,

1H), 4.33-4.31 (m, 1H), 4.09-4.08 (m, 1H), 2.93-2.88 (m, 1H), 2.68-2.64 (m, 1H). <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 150.26, 147.37, 146.82, 143.06, 128.43, 119.26, 119.07, 86.39, 86.33, 84.75, 71.17, 64.56, 64.53, 52.29, 39.26. <sup>31</sup>P{<sup>1</sup>H} (202 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 0.24. MS (ESI) Calcd. for  $C_{13}H_{19}N_5O_6P^+$ :  $[M]^+$ , 372.1; found: 372.2.

**N<sup>1</sup>-(*trans*-2-hexenyl)-2'-deoxyadenosine-5'-monophosphate (60e).** The reaction was done with 1.4

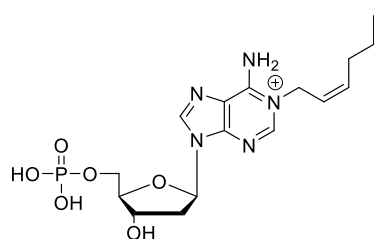


mmol nucleoside derivative and 3 eq *trans*-2-hexenyl bromide,

reaction time 48 h. Yield 8 mg (0.02 mmol, 14 %). <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 8.59 (s, 1H), 8.52 (s, 1H), 6.55 (t,  $J = 6.7$  Hz, 1H), 5.80-5.74 (m, 1H), 5.70-5.64 (m, 1H), 4.89-4.87 (m, 2H), 4.75-4.72 (m, 1H), 4.30-4.27 (m, 1H), 4.05 (dd,  $J = 5.2, 3.6$

Hz, 2H), 2.86 (ddd,  $J = 14.1, 6.9, 6.1$  Hz, 1H), 2.62 (ddd,  $J = 14.1, 6.5, 3.7$  Hz, 1H), 2.05 (qd,  $J = 7.0, 1.2$  Hz, 2H), 1.39-1.32 (m, 2H), 0.82 (t,  $J = 7.4$  Hz, 3H). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 150.11, 147.16, 146.76, 142.94, 137.67, 119.68, 119.15, 86.32, 86.25, 84.64, 71.14, 64.56, 64.52, 52.08, 39.23, 33.56, 21.20, 12.73. <sup>31</sup>P{<sup>1</sup>H} (202 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 0.14. MS (ESI) Calcd. for  $C_{16}H_{25}N_5O_6P^+$ :  $[M]^+$ , 414.2; found: 414.3.

**N<sup>1</sup>-(*cis*-2-hexenyl)-2'-deoxyadenosine-5'-monophosphate (60f).** The reaction was done with 1.4

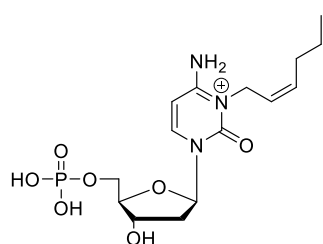


mmol nucleoside derivative and 4 eq *cis*-2-hexenyl bromide,

reaction time 48 h. Yield 10 mg (0.024 mmol, 16 %). <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 8.58 (s, 1H), 8.52 (s, 1H), 6.54 (t,  $J = 6.7$  Hz, 1H), 6.00-5.95 (m, 1H), 5.66-5.61 (m, 1H), 4.96-4.95 (m, 2H), 4.74-4.72 (m, 1H), 4.30-4.27 (m, 1H), 4.06-4.04 (m, 2H), 2.86 (ddd,

$J = 14.1, 6.9, 6.1$  Hz, 1H), 2.62 (ddd,  $J = 14.1, 6.6, 3.8$  Hz, 1H), 2.28-2.13 (m, 2H), 1.47-1.39 (m, 2H), 0.89 (t,  $J = 7.4$  Hz, 3H). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 150.36, 146.72, 146.68, 142.98, 139.14, 119.20, 118.64, 86.32, 86.25, 84.66, 71.13, 64.57, 64.53, 47.75, 39.22, 29.10, 21.69, 12.90. <sup>31</sup>P{<sup>1</sup>H} (202 MHz, D<sub>2</sub>O)  $\delta$ /ppm: 0.13. MS (ESI) Calcd. for  $C_{16}H_{25}N_5O_6P^+$ :  $[M]^+$ , 414.2; found: 414.3.

**N<sup>3</sup>-(*cis*-2-hexenyl)-2'-deoxycytidine-5'-monophosphate (61a).** The reaction was done with 1.4 mmol



nucleoside derivative and 2 eq *cis*-2-hexenyl bromide, reaction time 1 h.

Yield 2 mg (0.005 mmol, 3 %). <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ/ppm: 8.19

(d, *J* = 8.0 Hz, 1H), 6.32 (d, *J* = 8.0 Hz, 1H), 6.23 (t, *J* = 6.3 Hz), 5.84-

5.78 (m, 1H), 5.40-5.35 (m, 1H), 4.72-4.71 (m, 2H), 4.55-4.52 (m, 1H),

4.25-4.23 (m, 1H), 4.11 (ddd, *J* = 11.7, 4.3, 2.8 Hz), 4.03 (ddd, *J* = 11.7,

5.1, 3.0 Hz, 1H), 2.50 (ddd, *J* = 14.2, 6.3, 3.9 Hz, 1H), 2.36-2.31 (m, 1H), 2.19-2.14 (m, 2H), 1.45-

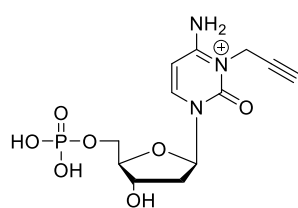
1.38 (m, 2H), 0.90 (t, *J* = 7.4 Hz, 3H). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ/ppm 158.96, 148.31, 141.90,

137.25, 118.75, 95.37, 87.79, 86.45, 70.54, 64.06, 41.72, 39.80, 29.16, 21.87, 12.99. <sup>31</sup>P{<sup>1</sup>H} (202

MHz, D<sub>2</sub>O) δ/ppm: 0.38. MS (ESI) Calcd. for C<sub>15</sub>H<sub>25</sub>N<sub>3</sub>O<sub>7</sub>P<sup>+</sup>: [M]<sup>+</sup>, 390.1; found: 390.2.

**Propargylation of 2'-deoxycytidine-5'-monophosphate.** The nucleoside derivative (1.5 mmol) was alkylated with 2.5 eq of propargyl bromide (as 80 % w/w solution in toluene) for 20 h at room temperature. Compounds **61b** and **61c** were isolated as the major products of the reaction.

**N<sup>3</sup>-(propargyl)-2'-deoxycytidine-5'-monophosphate (61b).** Yield 5 mg (0.014 mmol, 9 %). <sup>1</sup>H



NMR (500 MHz, D<sub>2</sub>O) δ/ppm: 8.18 (d, *J* = 8.0 Hz, 1H), 6.34 (d, *J* = 8.0

Hz, 1H), 6.25 (t, *J* = 6.3 Hz, 1H), 4.87 (t, *J* = 2.5 Hz, 2H), 4.55-4.53 (m,

1H), 4.26-4.24 (m, 1H), 4.12 (ddd, *J* = 11.8, 4.4, 2.7 Hz, 1H), 4.04 (ddd, *J*

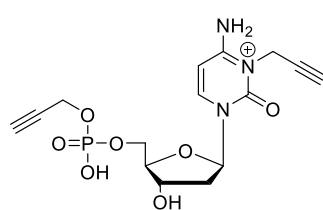
= 11.7, 5.1, 3.0 Hz, 1H), 2.79 (t, *J* = 2.5 Hz, 1H), 2.52 (ddd, *J* = 14.2, 6.3,

3.8 Hz, 1H), 2.37-2.32 (m, 1H). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ/ppm:

158.57, 147.86, 142.14, 95.38, 87.99, 86.49, 86.41, 75.09, 73.79, 70.56, 64.29, 39.78, 33.51. <sup>31</sup>P{<sup>1</sup>H}

(202 MHz, D<sub>2</sub>O) δ/ppm: -0.04. MS (ESI) Calcd. for C<sub>12</sub>H<sub>17</sub>N<sub>3</sub>O<sub>7</sub>P<sup>+</sup>: [M]<sup>+</sup>, 346.1; found: 346.1.

**N<sup>3</sup>-(propargyl)-2'-deoxycytidine-5'-(*O*-propargyl)phosphate (61c).** Yield 3 mg (0.008 mmol, 5 %).



<sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ/ppm: 8.15 (d, *J* = 8.0 Hz, 1H), 6.35 (d, *J* =

8.0 Hz, 1H), 6.24 (t, *J* = 6.4 Hz, 1H), 4.88 (dd, *J* = 2.6, 1.7 Hz, 2H),

4.58-4.55 (m, 1H), 4.51 (dd, *J* = 10.2, 2.5 Hz, 2H), 4.28-4.26 (m, 1H),

4.17 (ddd, *J* = 11.7, 4.3, 2.6 Hz, 1H), 4.09 (ddd, *J* = 11.7, 5.0, 3.1 Hz,

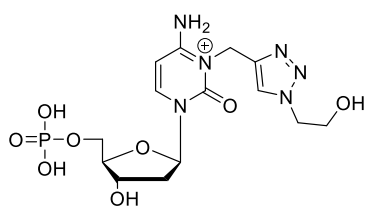
1H), 2.91 (t, *J* = 2.4 Hz, 1H), 2.80 (t, *J* = 2.5 Hz, 1H), 2.54 (ddd, *J* =

14.2, 6.3, 3.7 Hz, 1H), 2.38-2.33 (m, 1H). <sup>13</sup>C NMR (126 MHz, D<sub>2</sub>O) δ/ppm: 158.56, 147.83, 142.14,

95.34, 88.12, 86.40, 86.31, 75.80, 75.14, 73.79, 70.62, 64.80, 53.66, 53.60, 39.80, 33.53. <sup>31</sup>P{<sup>1</sup>H} (202

MHz, D<sub>2</sub>O) δ/ppm: -0.02. MS (ESI) Calcd. for C<sub>15</sub>H<sub>19</sub>N<sub>3</sub>O<sub>7</sub>P<sup>+</sup>: [M]<sup>+</sup>, 384.1; found: 384.2.

**N<sup>3</sup>-((2-Hydroxyethyl)-1*H*-1,2,3-triazol-4-yl)methyl-2'-deoxycytidine-5'-monophosphate (61d).** An



NMR sample from compound **61b** in D<sub>2</sub>O (600 μl) was mixed with 2-azidoethan-1-ol (5 μl), 12.5 mM CuSO<sub>4</sub> in D<sub>2</sub>O (24 μl), and 100 mM sodium ascorbate in D<sub>2</sub>O (30 μl). <sup>1</sup>H NMR after 5 min at room temperature showed complete disappearance of the resonance for the terminal alkyne proton. The whole mixture was purified by

preparative HPLC. <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ/ppm: 8.21 (d, *J* = 8.0 Hz, 1H), 6.36 (d, *J* = 8.0 Hz, 1H), 6.24 (t, *J* = 6.3 Hz, 1H), 5.35 (s, 2H), 4.54-4.52 (m, 3H), 4.25-4.23 (m, 1H), 4.11 (ddd, *J* = 11.7, 4.2, 2.6 Hz, 1H), 4.03 (ddd, *J* = 11.7, 5.2, 3.1 Hz, 1H), 3.98-3.96 (m, 2H), 2.51 (ddd, *J* = 14.2, 6.3, 3.9 Hz, 1H), 2.37-2.32 (m, 1H). <sup>13</sup>C NMR (determined from <sup>1</sup>H-<sup>13</sup>C HMQC and <sup>1</sup>H-<sup>13</sup>C HMBC, D<sub>2</sub>O) δ/ppm: 159.14, 148.18, 142.30, 139.49, 125.40, 95.29, 87.88, 86.41, 70.50, 64.16, 59.94, 52.52, 39.68, 39.00. <sup>31</sup>P{<sup>1</sup>H} (202 MHz, D<sub>2</sub>O) δ/ppm: 0.17. MS (ESI) Calcd. for C<sub>14</sub>H<sub>22</sub>N<sub>6</sub>O<sub>8</sub>P<sup>+</sup>: [M+H]<sup>+</sup>, 433.1; found: 434.2.

#### 5.4. Oligonucleotide synthesis and purification

Solid-phase oligonucleotide synthesis was carried out on 1-μmol CPG columns using standard phosphoramidite chemistry with 0.3 M 5-benzylthio-1*H*-tetrazole as activator. The DNA oligonucleotides were cleaved from the support with 32 % (v/v) aqueous ammonia for 2 h at room temperature and deprotected for 18 h at 55 °C, then freeze-dried and purified by micropreparative HPLC using method A or by semi-preparative HPLC.

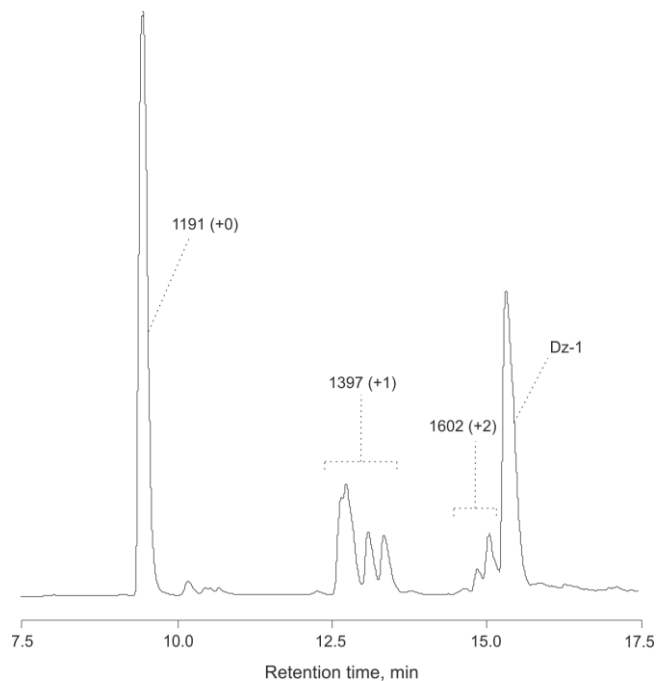
The RNA oligonucleotides were synthesized using TBDMS protection strategy for the 2'-OH group. Cleavage from the support was done in 32 % aqueous ammonia/ethanol 3:1 (v/v) for 2 h at room temperature followed by deprotection for 12 h at 55 °C. The samples were freeze-dried and the TBDMS groups were cleaved with triethylamine/triethylamine trihydrofluoride/*N*-methylpyrrolidine 1.5:2:3 (v/v/v) for 2 h at 65 °C. The crude oligoribonucleotides were isolated by precipitation with 3 M NaOAc (1/10 volume) and *n*-butanol (4 volumes) for 2 h at -78 °C followed by centrifugation (15,000 × g, 10 min). The pellets were washed with 70 % (v/v) ethanol and dried under high vacuum. The samples were further purified by micropreparative HPLC using method A.

The identity of all synthesized oligonucleotides was confirmed by ESI MS or MALDI TOF.

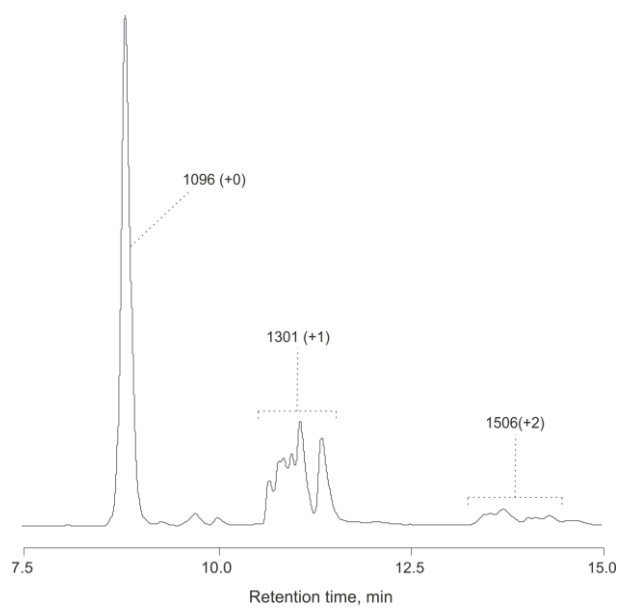
#### 5.5. Nucleic acid alkylation with rhodium(II) carbenoids

**General procedure for Rh<sub>2</sub>(OAc)<sub>4</sub>-catalyzed oligonucleotide modification using α-diazocarbonyl compounds.** Typically 10 or 20 μl reaction mixtures containing 5 mM oligonucleotide, 500 μM Rh<sub>2</sub>(OAc)<sub>4</sub> and 50 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at 20

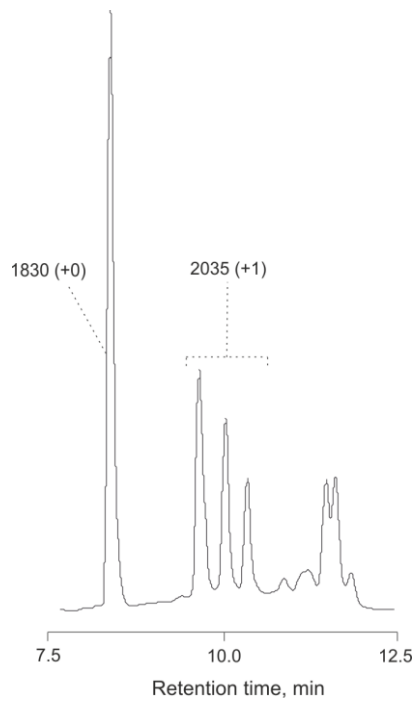
°C for 24-72 h. Analysis of the reaction products was carried out by micropreparative HPLC-separation of 5- $\mu$ l aliquots of the reaction mixtures using method A. The collected peak fractions were further assessed by ESI-MS or MALDI TOF.



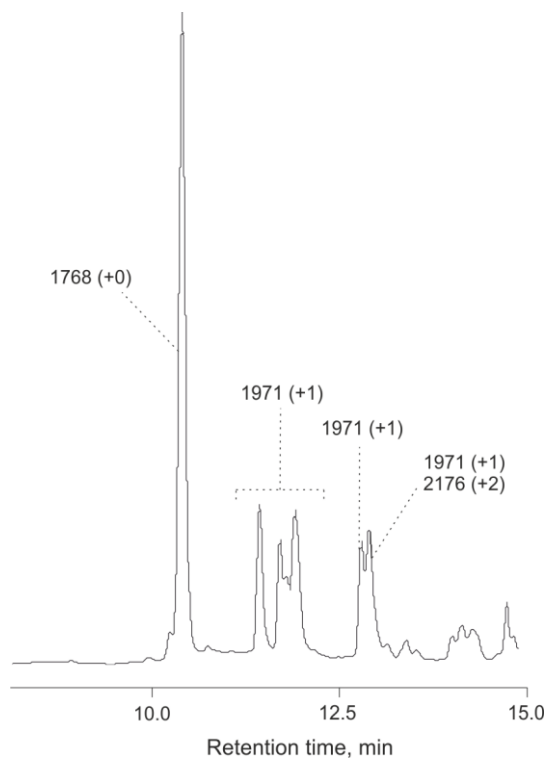
**Figure E3** | HPLC trace of the modification reaction of d(A<sub>4</sub>) with Dz1 (HPLC analysis by method A).



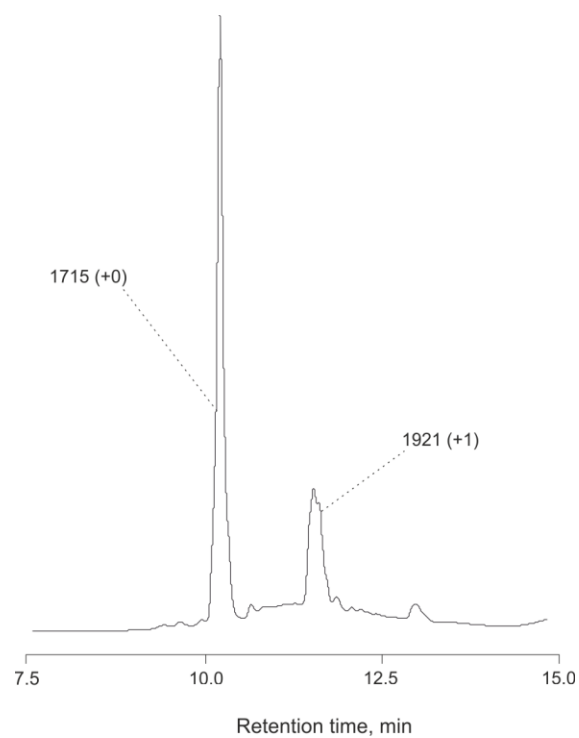
**Figure E4** | HPLC trace of the modification reaction of d(C<sub>4</sub>) with Dz1 (HPLC analysis by method A).



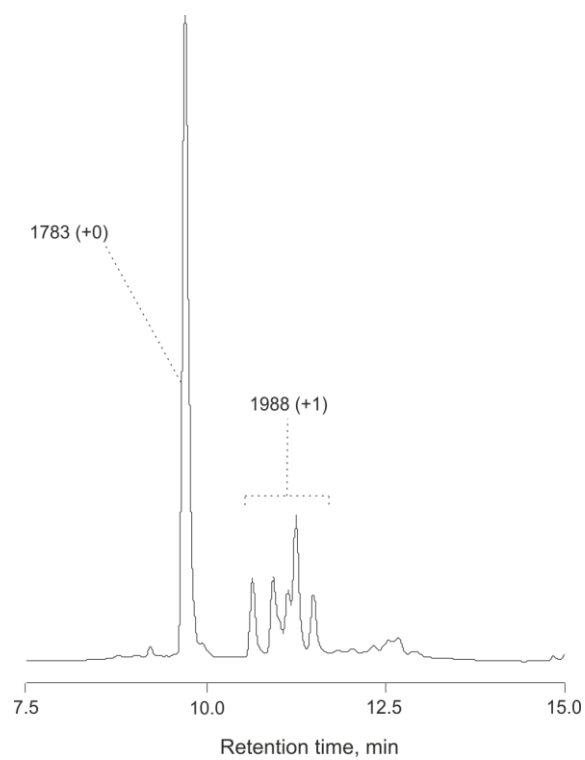
**Figure E5** | HPLC trace of the modification reaction of a mixture of r(ACU GCU) with Dz1 (HPLC analysis by method A).



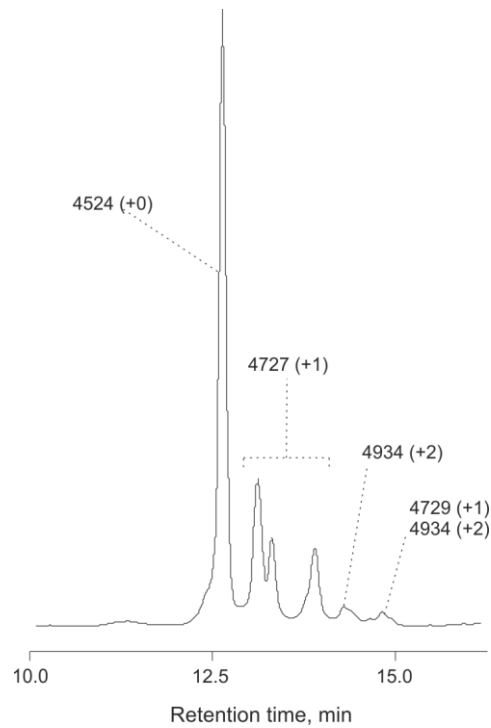
**Figure E6** | HPLC trace of the modification reaction of d(ACT GCT) with Dz1 (HPLC analysis by method A).



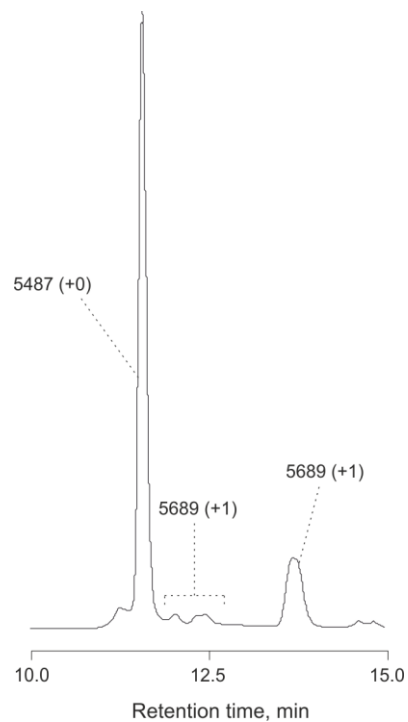
**Figure E7** | HPLC trace of the modification reaction of d(CTC TCT) with Dz1 (HPLC analysis by method A).



**Figure E8** | HPLC trace of the modification reaction of d(CTG GCT) with Dz1 (HPLC analysis by method A).

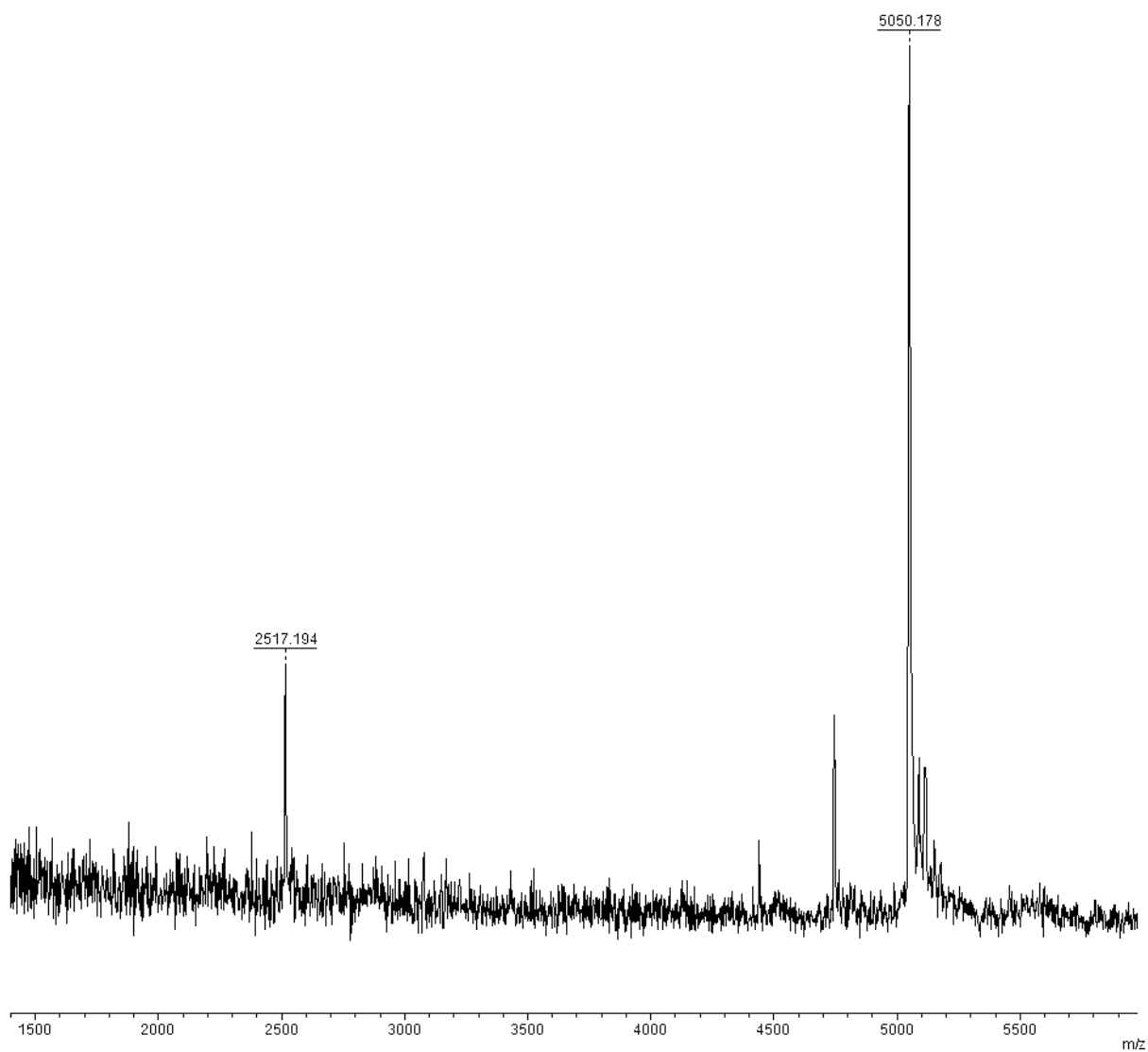


**Figure E10** | HPLC trace of the modification reaction of d(TTT ATT TGT TTC TTT) with Dz1 (HPLC analysis by method A).

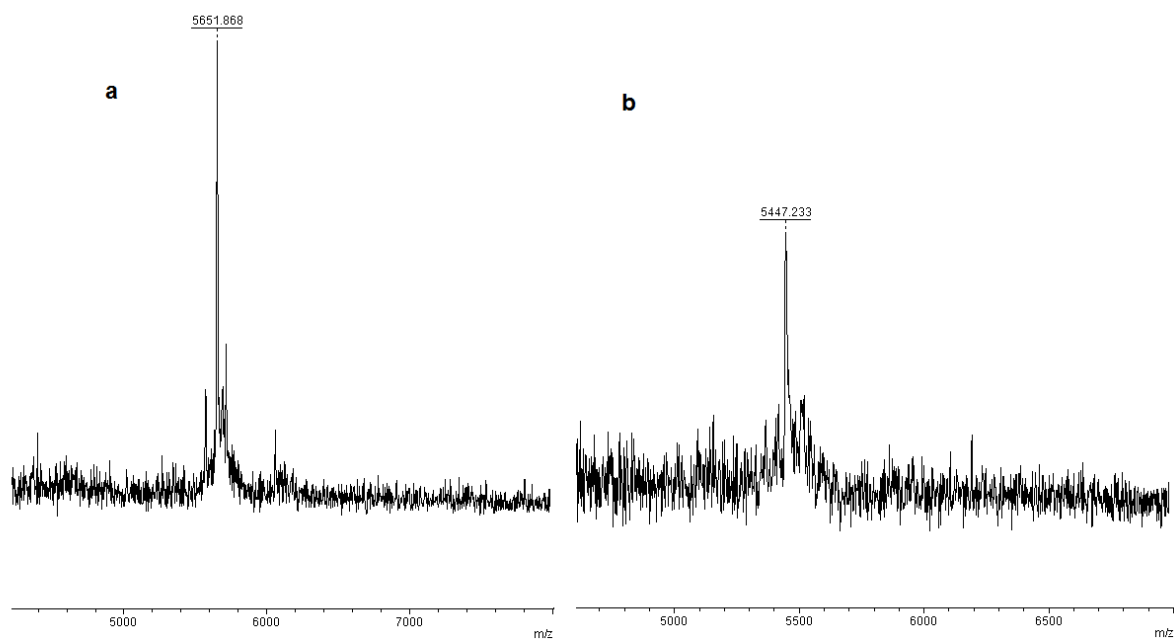


**Figure E11** | HPLC trace of the modification reaction of d(CGA ACG TTT TTC GTT CGA) (hairpin DNA oligonucleotide with a 3'-adenine overhang) with Dz1 (HPLC analysis by method A).

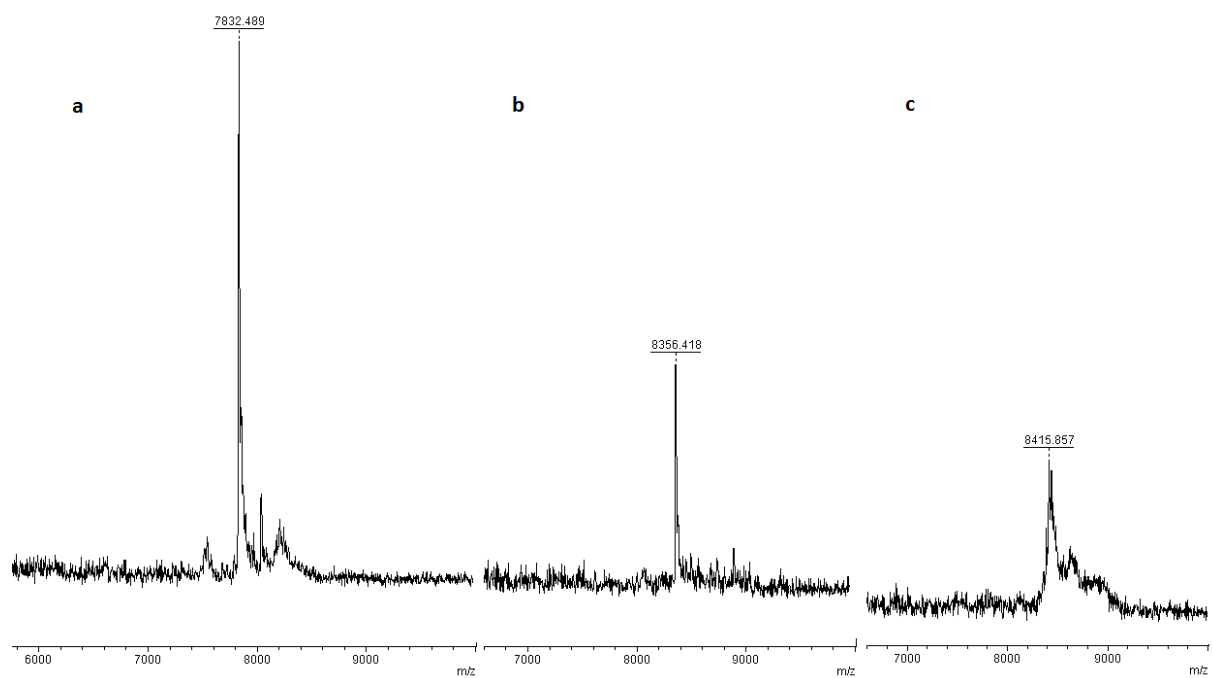




**Figure E12** | MALDI TOF MS spectrum of the main peak fraction of the modification reaction of d(TTT ATT TGT TTC TTT) with Dz5.



**Figure E13** | MALDI TOF MS spectrum of the ‘click’ products of the propargylated d(TTT ATT TGT TTC TTT) with (a) Rhodamine B azide and (b) biotin azide.



**Figure E14** | MALDI TOF MS of the singly (a) and doubly (b) propargylated d(C4)-T7 promoter primer and the ‘click’ reaction products of the total propargylated fraction with Rhodamine B azide (c).

### 5.5.1. NMR Characterization of d(TCT), d(TAT) and d(TGT) and the products of alkylation with Dz1

All measurements were performed on a Bruker Avance III NMR spectrometer (600.13 MHz proton frequency) equipped with a 5-mm BBFO smart probe with shielded  $z$ -gradient coil. All measurements were carried out at 298 K in Shigemi Microcell NMR tubes (280  $\mu$ l sample volume).

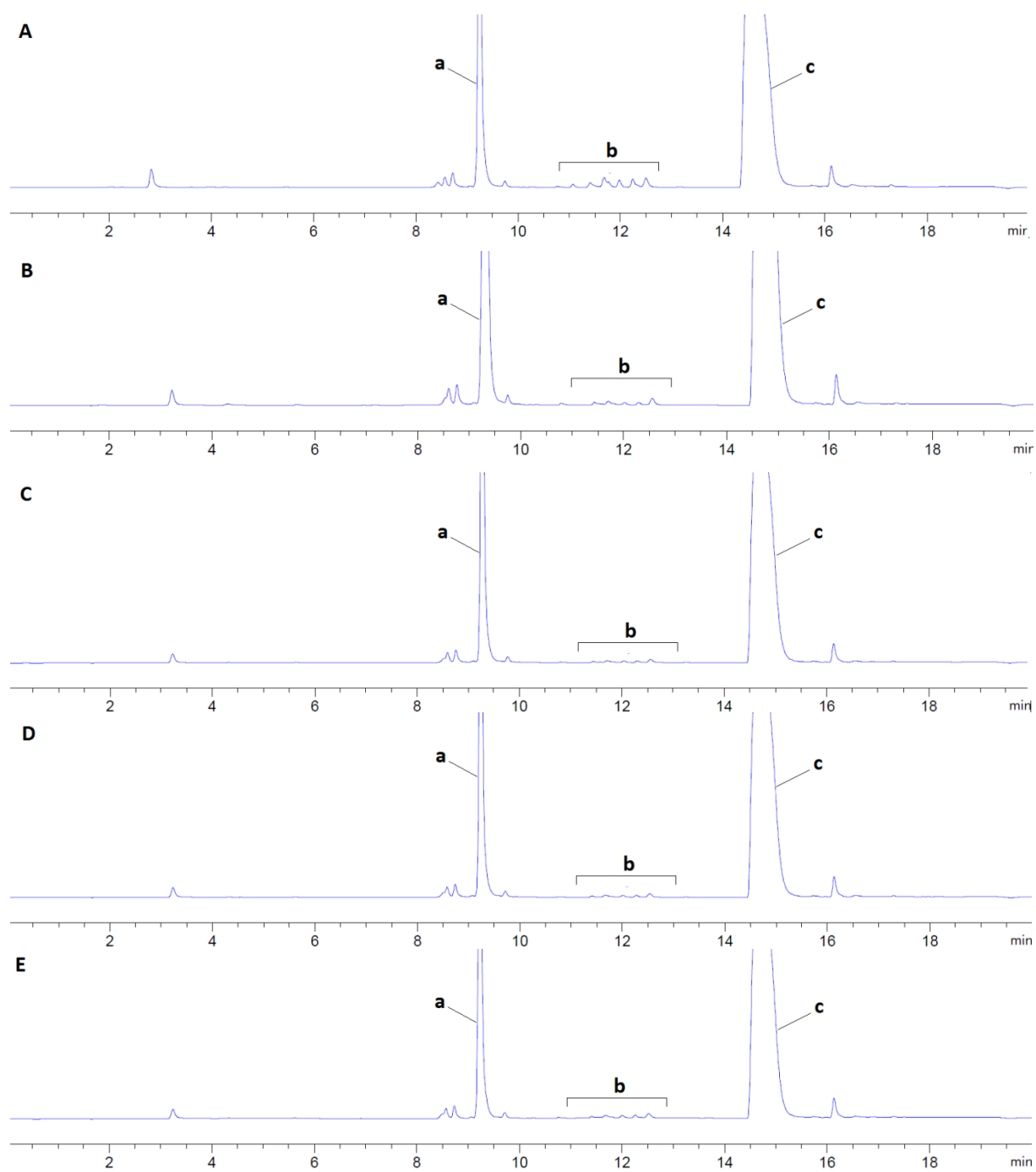
The HPLC-purified oligonucleotide was dissolved in 300  $\mu$ l of 10 mM potassium phosphate buffer, pH 6.0, containing 5 % (v/v) D<sub>2</sub>O. The pH was then adjusted to 6.00 with dilute HCl and NaOH. For water suppression, an excitation sculpting pulse sequence was used.<sup>103</sup> For experiments in D<sub>2</sub>O, the sample was freeze-dried from the H<sub>2</sub>O-containing phosphate buffer, re-dissolved in 100  $\mu$ l of D<sub>2</sub>O, freeze-dried again and then dissolved in D<sub>2</sub>O (99.99 %, Aldrich). It was then used directly without further adjustment of the pH.

The unambiguous assignment of the proton, carbon and phosphorous resonances was performed using standard COSY, long range - COSY, TOCSY, ROESY, HMQC, <sup>1</sup>H-<sup>13</sup>C HMBC and <sup>1</sup>H-<sup>31</sup>P-HMBC experiments.

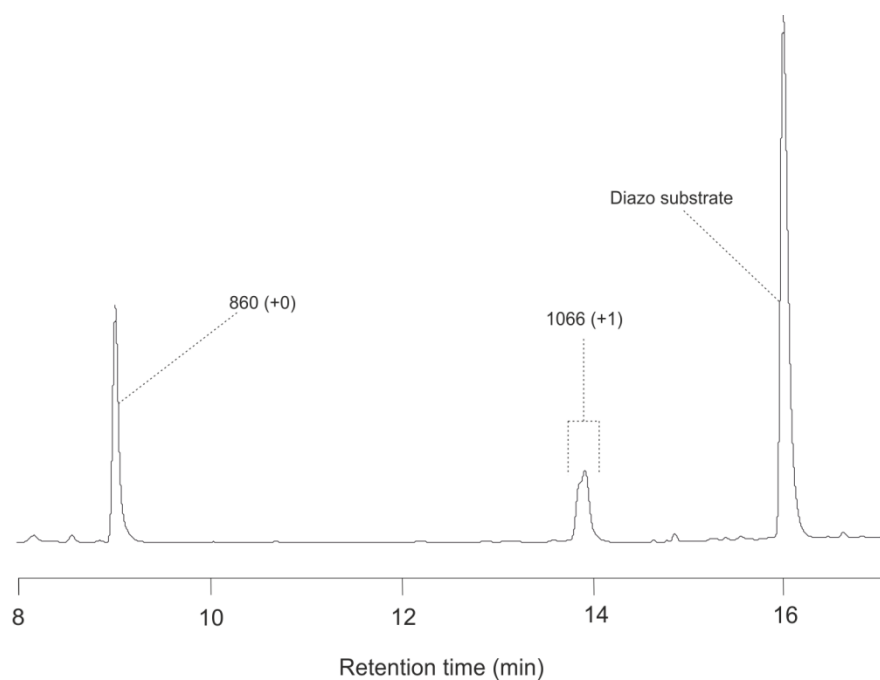
### 5.6. Nucleic acid alkylation with copper(I) carbenoids

**General procedure for Cu(I)-catalyzed oligonucleotide modification using  $\alpha$ -diazocarbonyl compounds.** Typically 10 or 20  $\mu$ l reaction mixtures containing 5 mM oligonucleotide, 500  $\mu$ M CuSO<sub>4</sub>, 2.5 mM THPTA and 50 mM  $\alpha$ -diazocarbonyl compound, and 10 mM sodium ascorbate in 100 mM MES buffer, pH 6.0 were kept at 20 °C for 24-72 h. Analysis of the reaction products was carried out by micropreparative HPLC-separation of 5- $\mu$ l aliquots of the reaction mixtures using methods A or C. The analysis of the modification reactions of d(ATG C) and d(ATG) was done in two different buffer systems (methods C and D) in order to fully resolve some of the modification products from the unreacted diazo substrate. The collected peak fractions were further assessed by ESI-MS or MALDI TOF. The designations of the chromatographic peak fractions (Supplementary Figures S1, 3, 6, 8-16) indicate the molecular mass of the species, and the number of alkylations per molecule of oligonucleotide (in parenthesis)

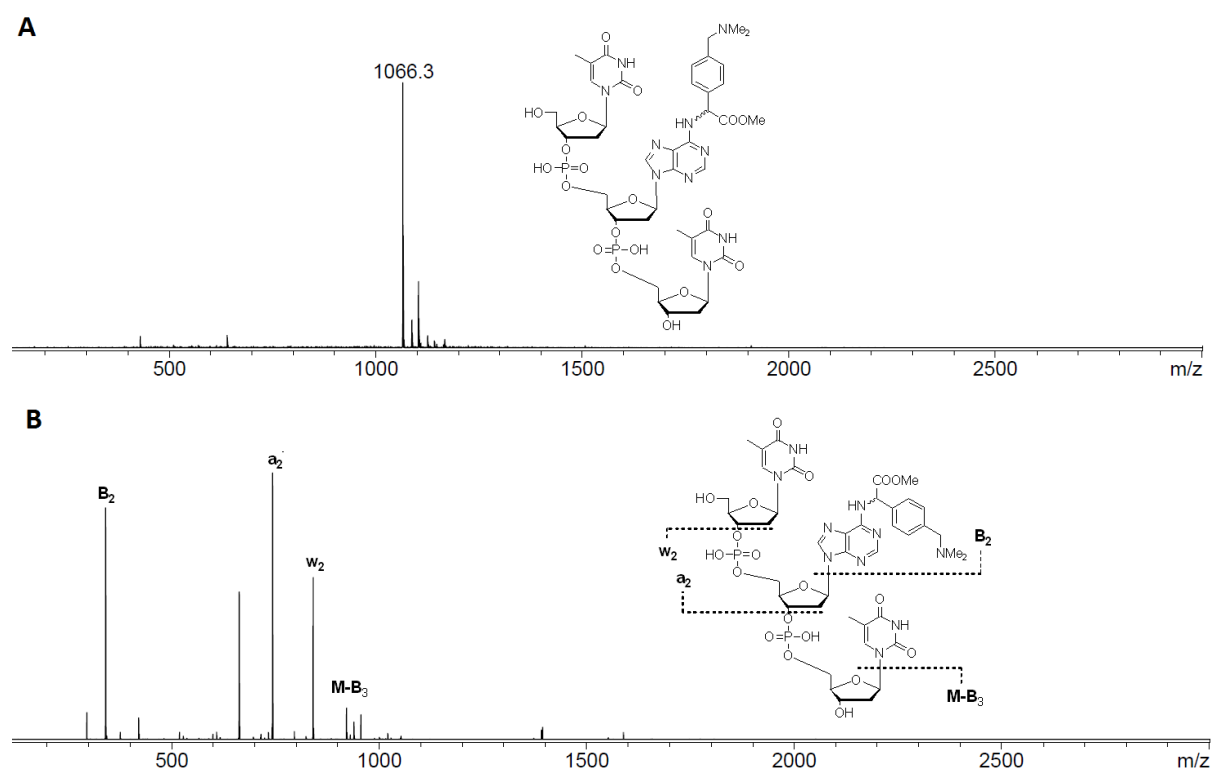
The singly charged ions of the modified products of the three trimers d(TAT), d(TGT) and d(TCT) from the ESI-MS were subjected to tandem mass spectrometry.



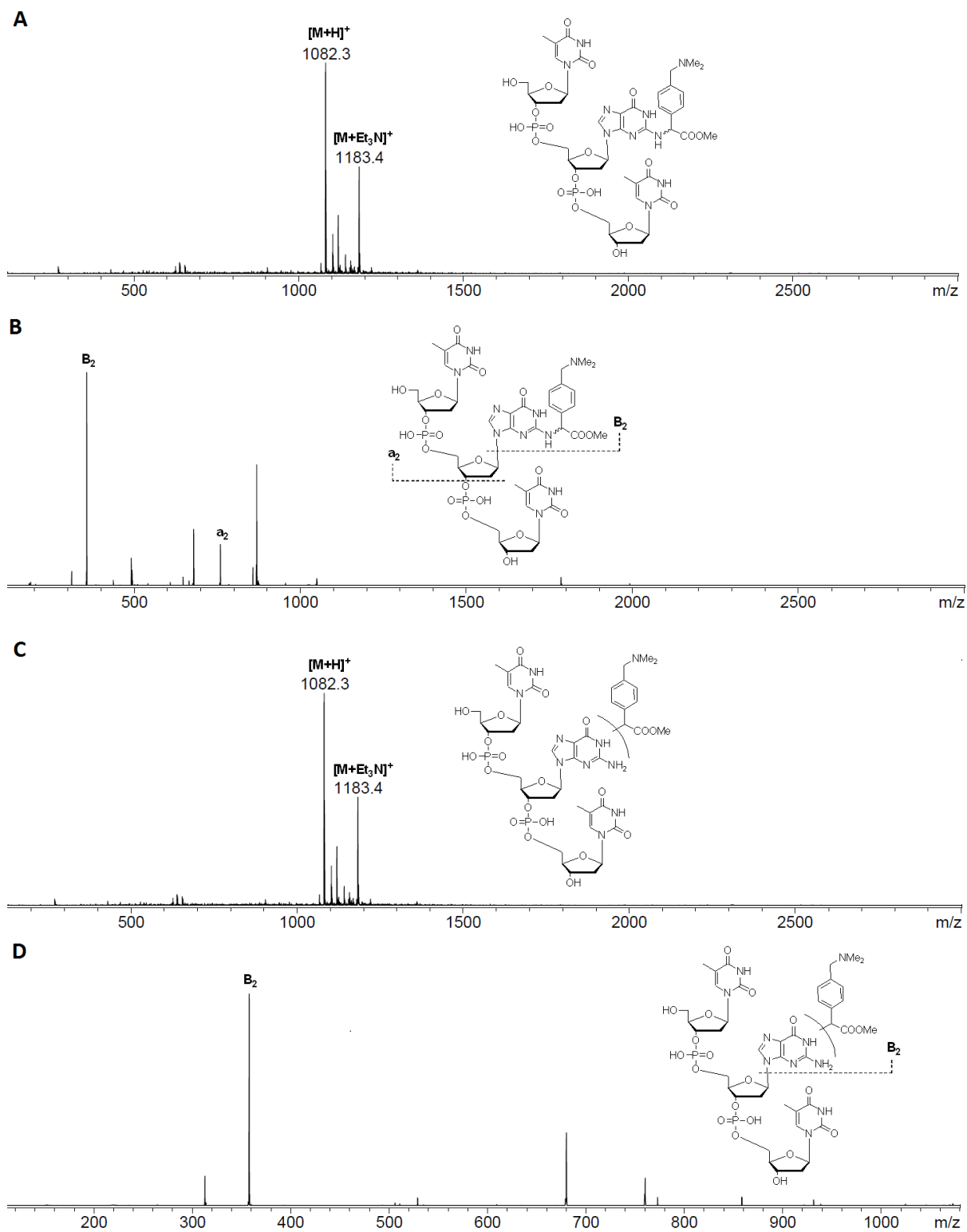
**Figure E15** | HPLC traces of the screening for metal catalyst for ON alkylation using method C (see Table 4). (a) Starting oligonucleotide (b) Modified products (c) Non-converted diazo substrate.



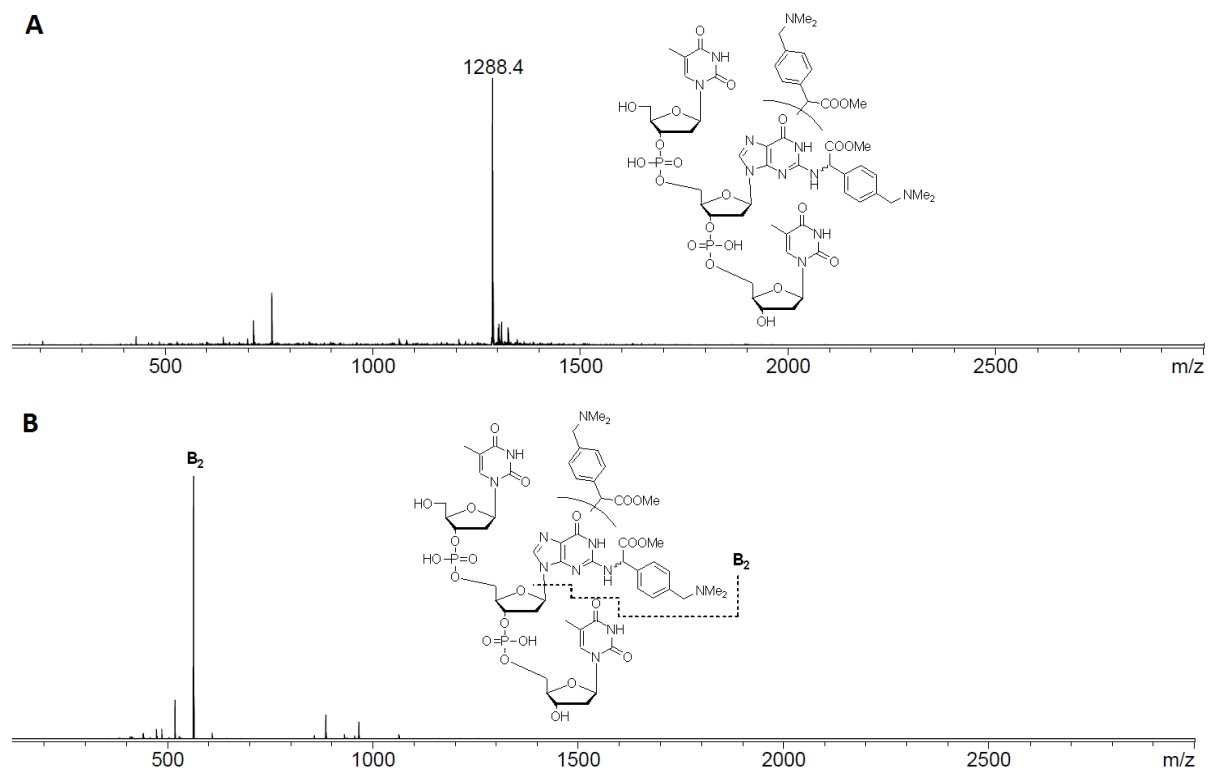
**Figure E16** | HPLC trace of the modification reaction of a d(TAT) with Dz1 (HPLC analysis by method C). The peak labels refer to the masses observed in the ESI-MS for those peaks. The (+1), (+2), or (+3) label indicates singly-, doubly-, or triply alkylated products.



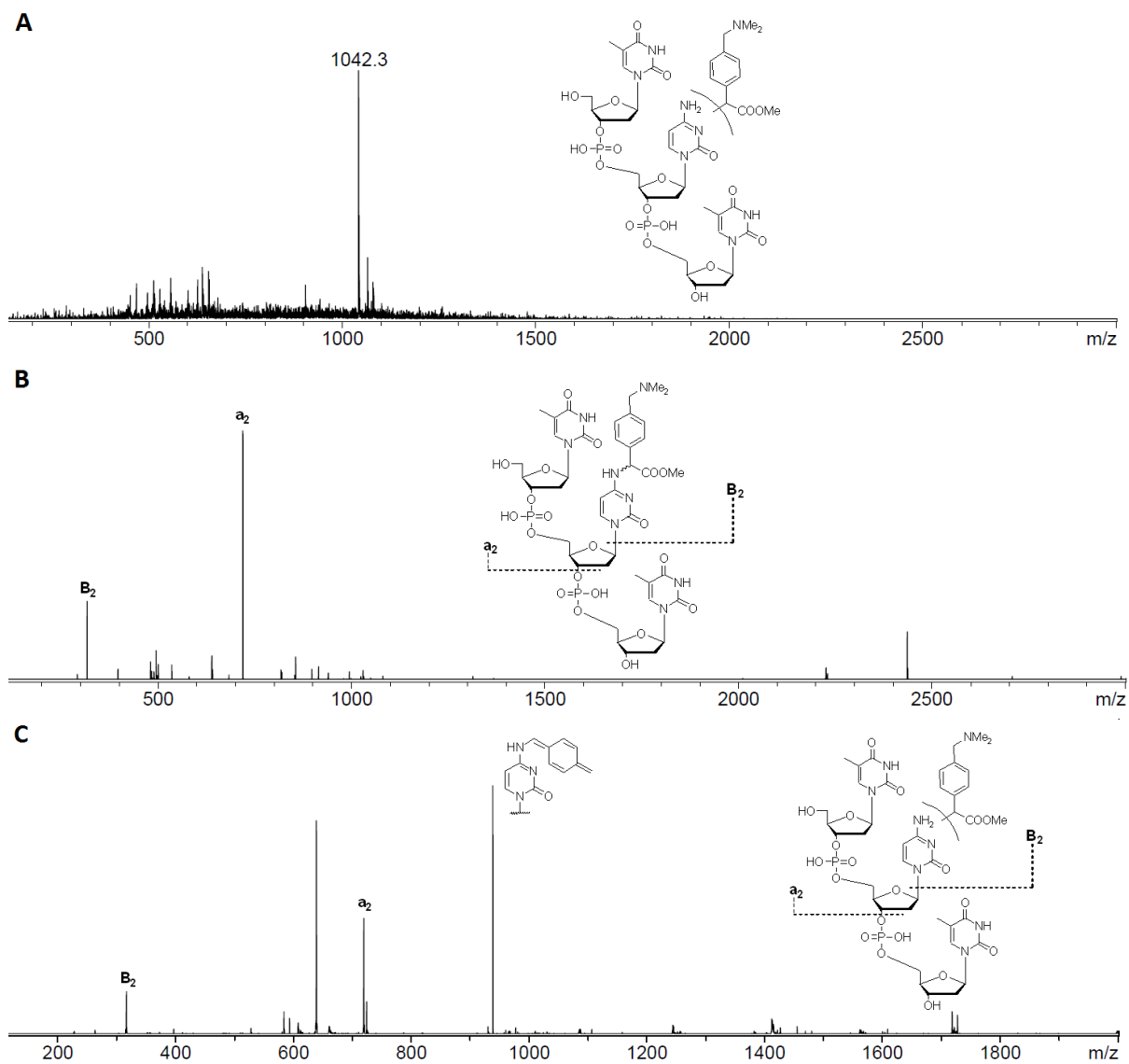
**Figure E17** | ESI-MS (A) and MS/MS analysis (B) of the singly-modified d(TAT) with Dz1.



**Figure E18** | ESI-MS (A, C) and MS/MS analysis (B, D) of the singly-modified d(TGT) with Dz1.

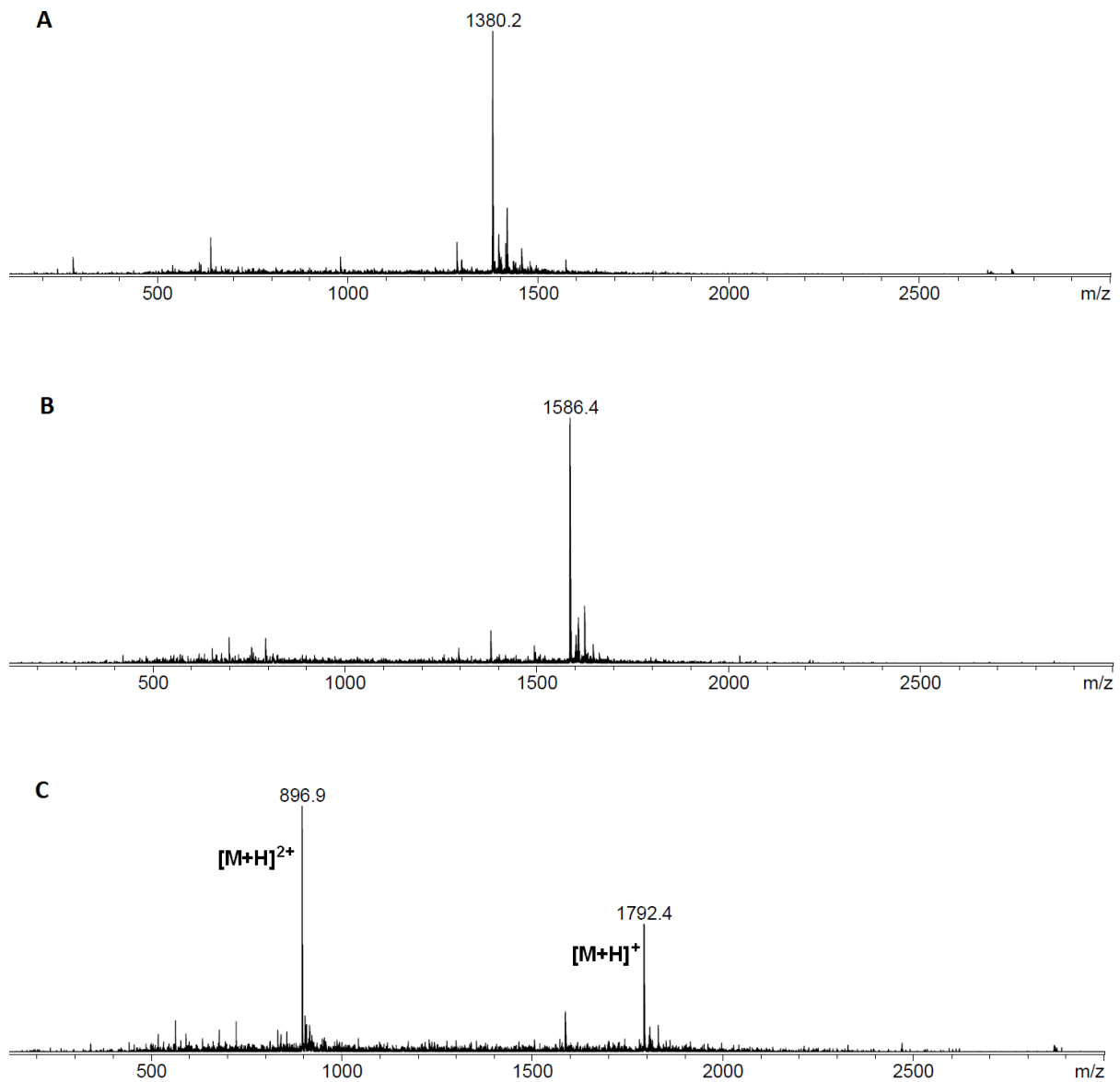


**Figure E19** | ESI-MS (A) and MS/MS analysis (B) of the doubly-modified d(TGT) with Dz1.

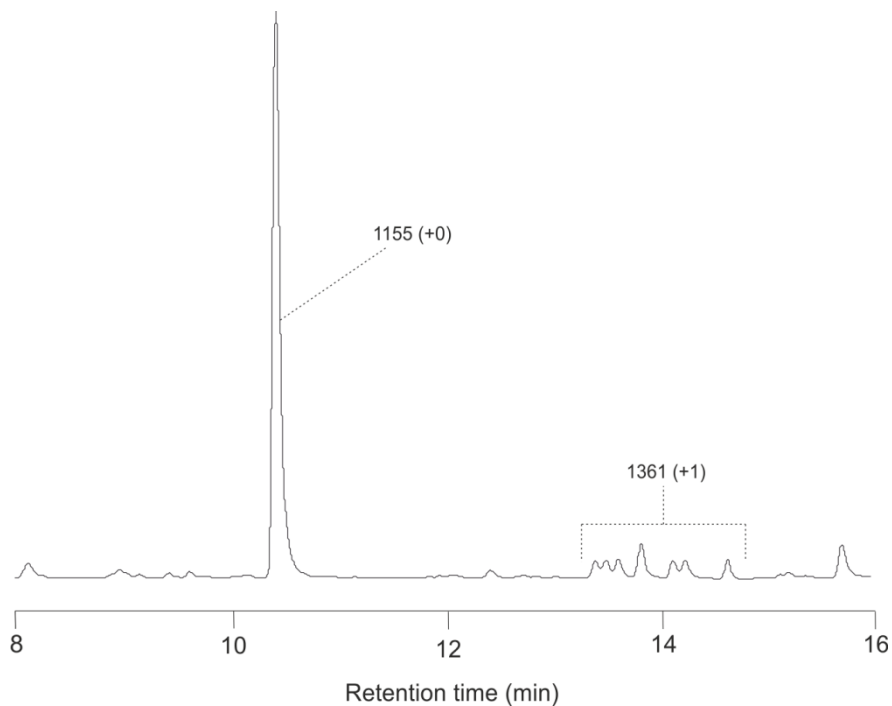


**Figure E20** | ESI-MS (A) and MS/MS analysis (B, C) of the singly-modified d(TCT) with Dz1.

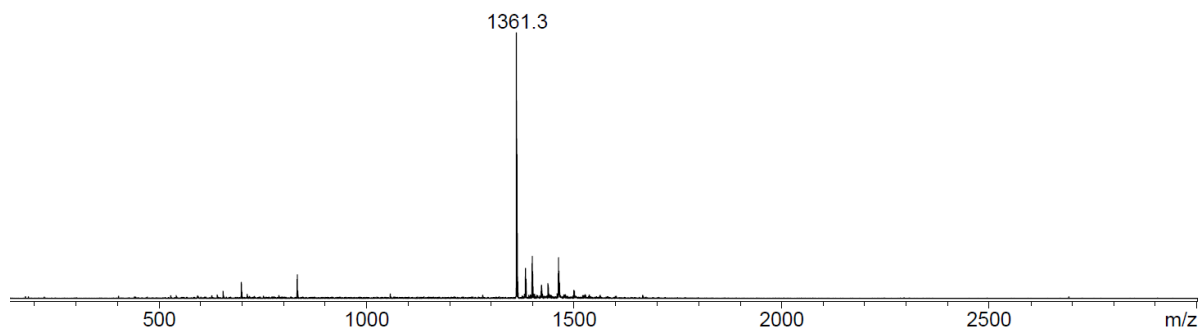




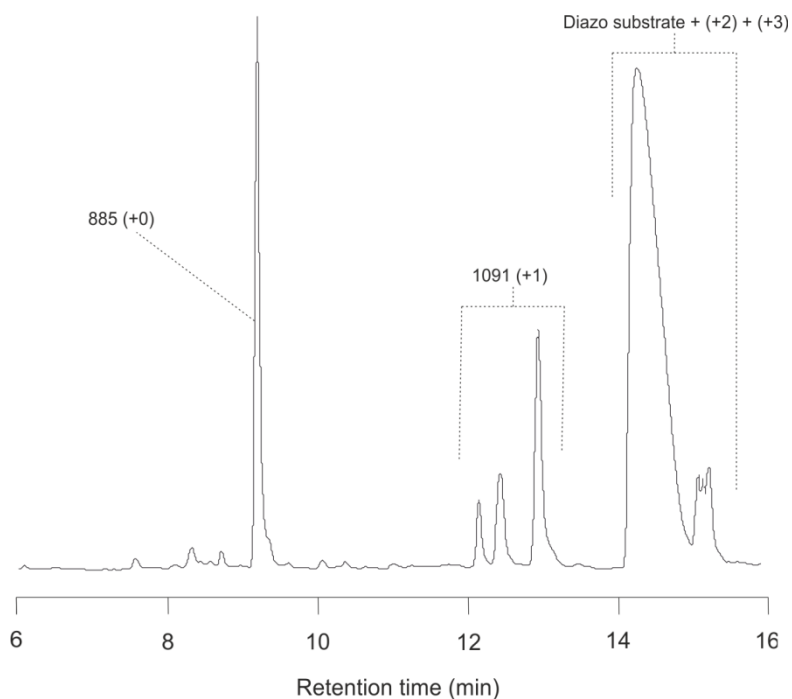
**Figure E21** | ESI-MS of singly (A), doubly (B) and triply modified (C) d(ATG C) with Dz1.



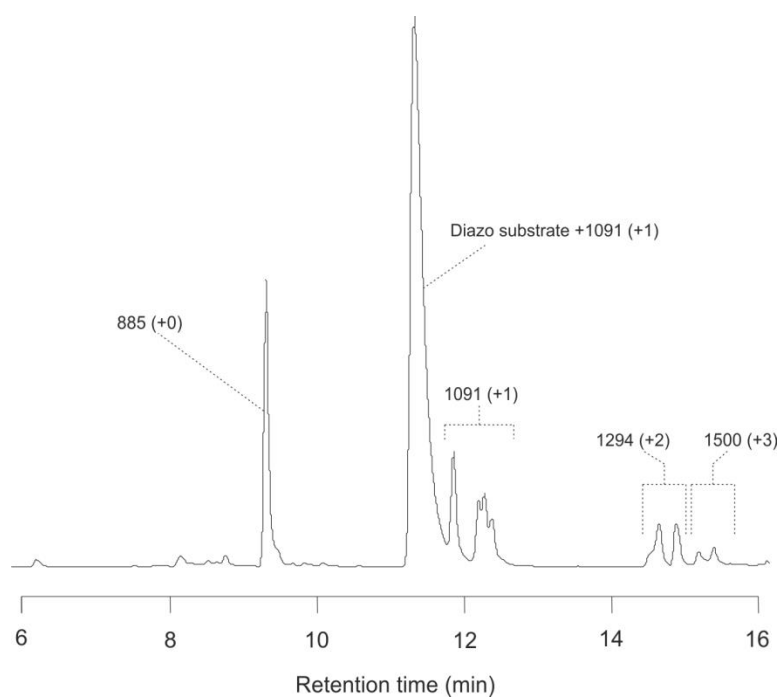
**Figure E22** | HPLC trace of the modification reaction of d(T<sub>4</sub>) with Dz1 (HPLC analysis by method C). The peak labels refer to the masses observed in the ESI-MS for those peaks. The (+1), (+2), or (+3) label indicates singly-, doubly-, or triply alkylated products.



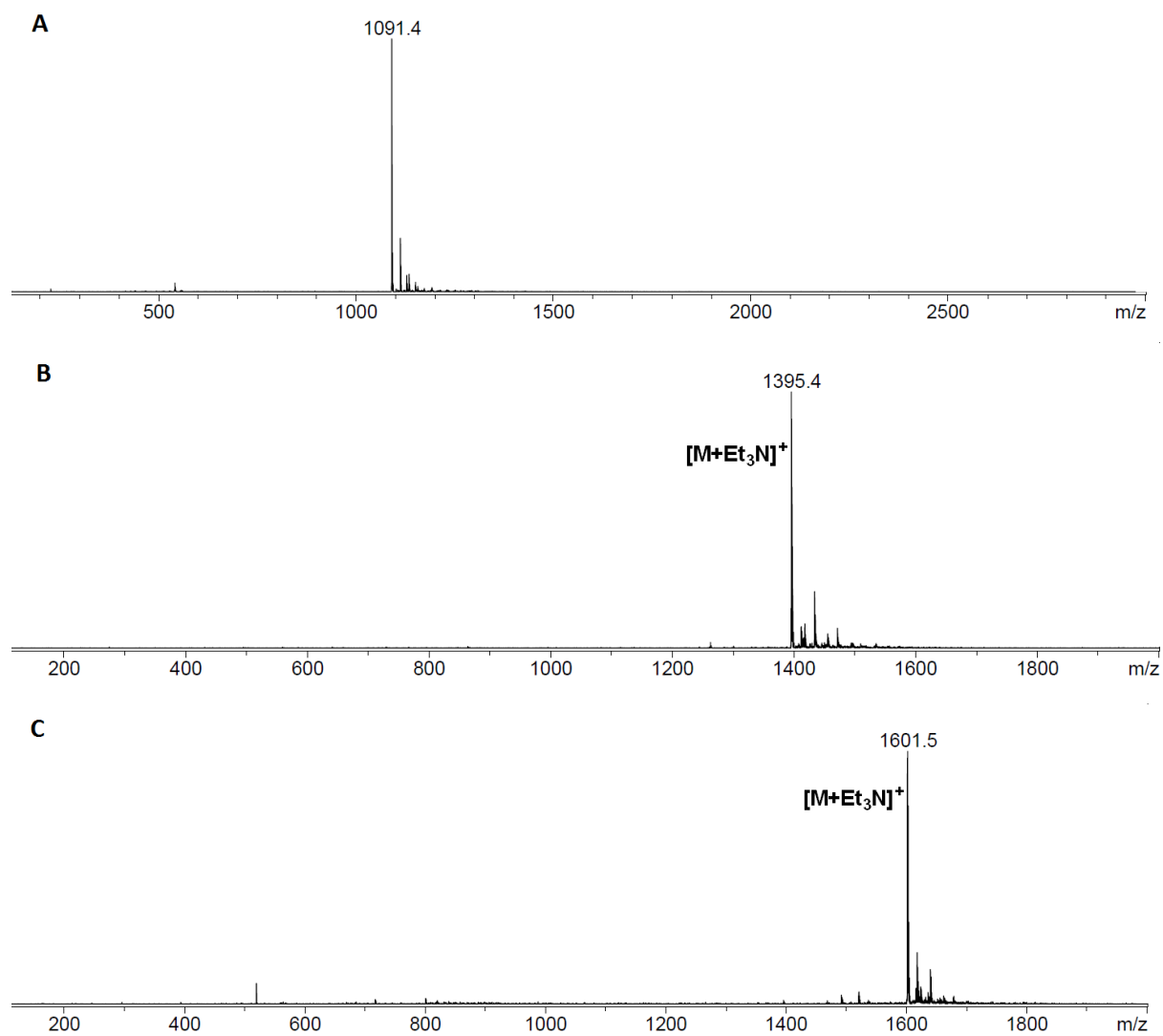
**Figure E23** | ESI-MS of singly modified d(T<sub>4</sub>) with Dz1.



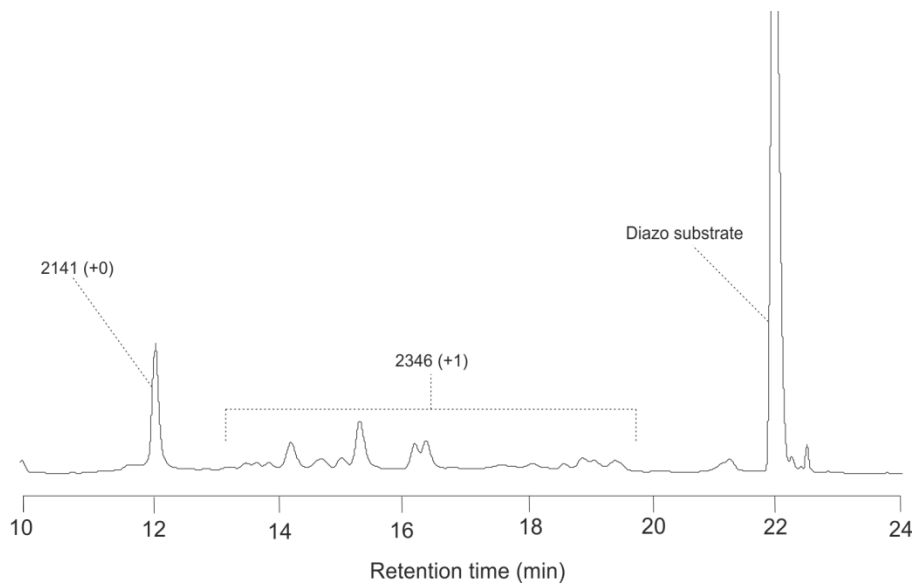
**Figure E24** | HPLC trace of the modification reaction of d(ATG) with Dz1 (HPLC analysis by method C). The peak labels refer to the masses observed in the ESI-MS for those peaks. The (+1), (+2), or (+3) label indicates singly-, doubly-, or triply alkylated products.



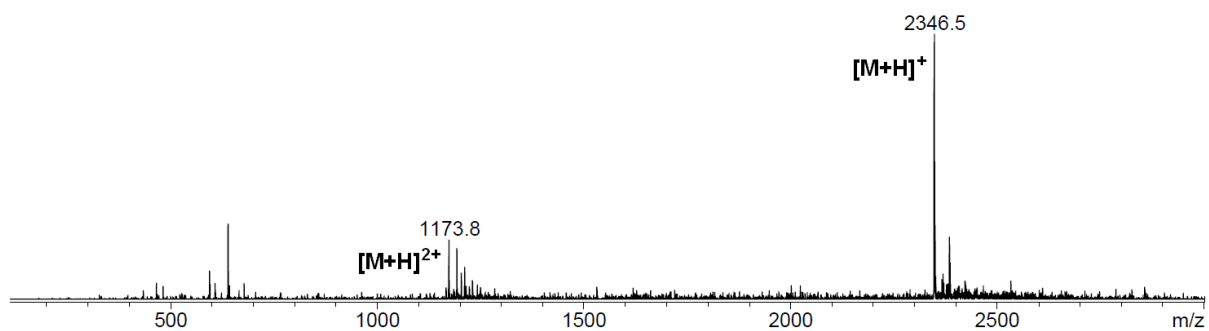
**Figure E25** | HPLC trace of the modification reaction of a d(ATG) with Dz1 (HPLC analysis by method D). The peak labels refer to the masses observed in the ESI-MS for those peaks. The (+1), (+2), or (+3) label indicates singly-, doubly-, or triply alkylated products.



**Figure E26** | ESI-MS of singly (A), doubly (B) and triply modified (C) d(ATG) with Dz1. The doubly and triply modified species were observed as triethylammonium adducts.



**Figure E27** | HPLC trace of the modification reaction of r(ACU GCU C) with Dz1 (HPLC analysis by method E). The peak labels refer to the masses observed in the ESI-MS for those peaks. The (+1), (+2), or (+3) label indicates singly-, doubly-, or triply alkylated products.



**Figure E28** | ESI-MS of singly modified r(ACU GCU C) with Dz-1.

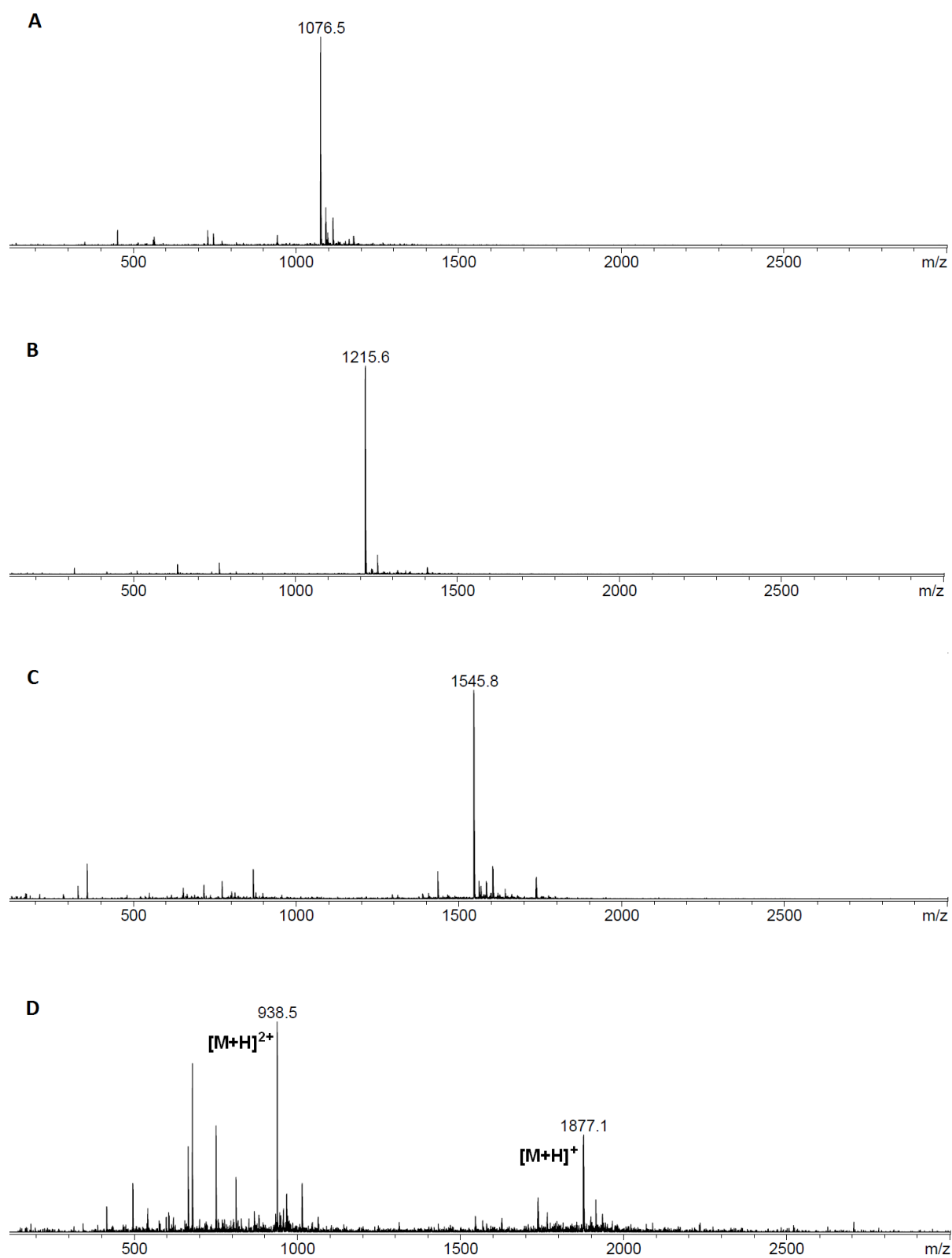
### **5.6.1. Cu(I)-catalyzed oligonucleotide modification using $\alpha$ -diazocarbonyl compounds in the presence of proteins.**

Protein concentration was determined from the absorbance at 279 or 280 nm using the following extinction coefficients: 43 824 M<sup>-1</sup>cm<sup>-1</sup> for bovine serum albumin,<sup>104</sup> 38 940 M<sup>-1</sup>cm<sup>-1</sup> for lysozyme,<sup>105</sup> and 41 820 M<sup>-1</sup>cm<sup>-1</sup> for streptavidine.<sup>106</sup> Typically 20  $\mu$ l reaction mixtures containing 5 mM d(TAT), 1 mM protein of interest, 500  $\mu$ M CuSO<sub>4</sub>, 2.5 mM THPTA, 50 mM  $\alpha$ -diazocarbonyl compound, and 10 mM sodium ascorbate in 100 mM MES buffer, pH 6.0 were kept at 20 °C for 24 h. Analysis of the reaction products was carried out by HPLC using method A. Aliquots (5- $\mu$ l) of the reaction mixtures were then separated micropreparatively and the collected peak fractions were further assessed by ESI-MS. The effect of the protein on the modification reaction was estimated by comparison of the conversion of the oligonucleotide and the diazo substrate in the presence of protein in a control sample.

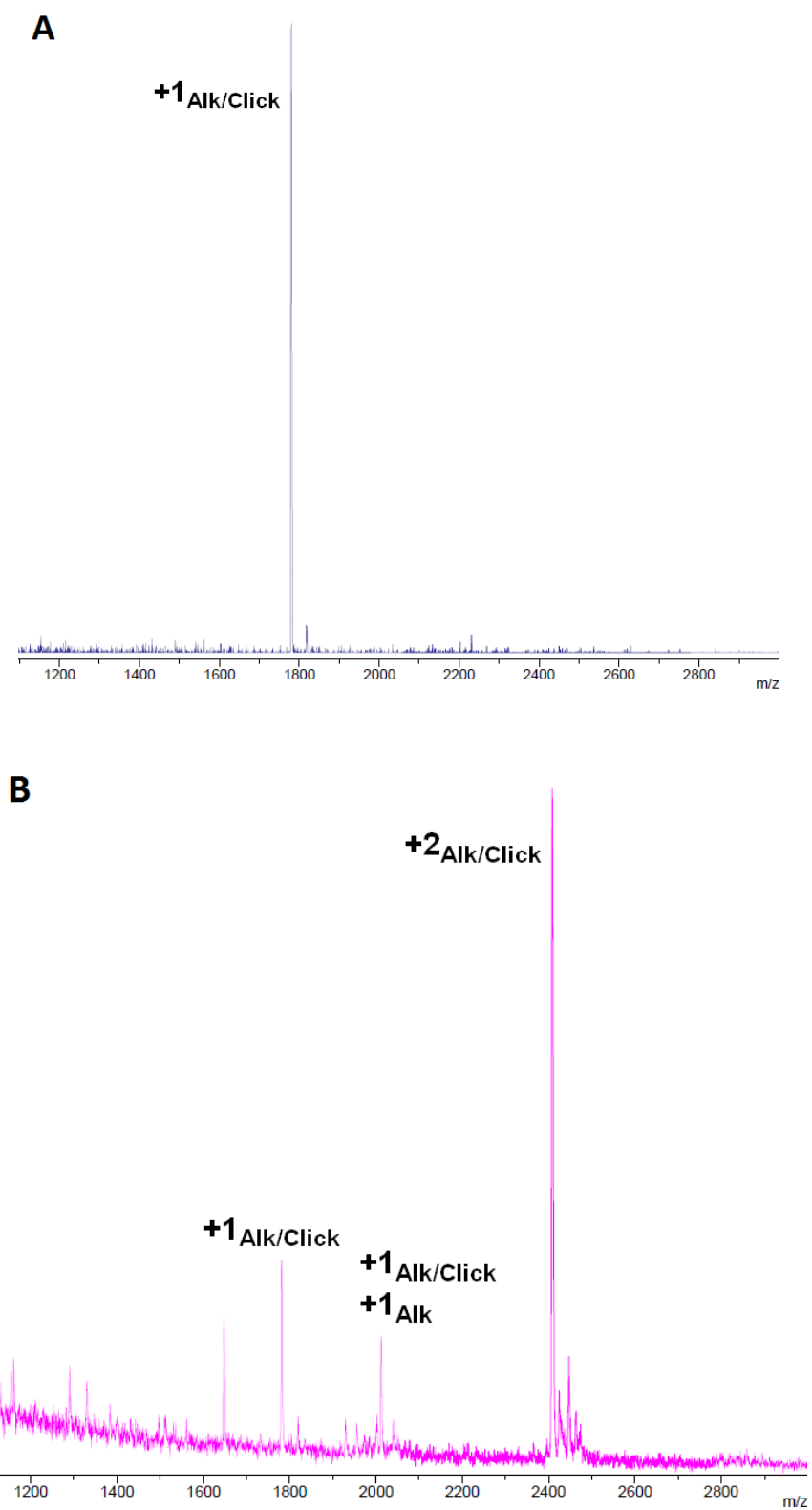
### **5.6.2. Auto-catalytic tandem CuAAC/N-H insertion with nucleic acids**

**Tandem alkylation/click reaction using diazo compound 57 and 3-azidoprop-1-ol.** Typically 10 or 20  $\mu$ l reaction mixtures containing 5 mM oligonucleotide, 500  $\mu$ M CuSO<sub>4</sub>, 2.5 mM THPTA, 50 mM compound 57, 55 mM of the azide and 10 mM sodium ascorbate in 100 mM MES buffer, pH 6.0 were kept at 20 °C for 24 h. Analysis of the reaction products was carried out by micropreparative HPLC-separation of 5- $\mu$ l aliquots of the reaction mixtures using method C. The collected peak fractions were further assessed by ESI-MS or MALDI TOF.

**Tandem alkylation/click reaction using diazo compound 57 and 1-(2-azidoacetyl)-4-(D-biotinyl)piperazine) 47.** Typically 10 or 20  $\mu$ l reaction mixtures containing 2.5 mM oligonucleotide, 250  $\mu$ M CuSO<sub>4</sub>, 1.25 mM THPTA, 25 mM propargyl-tagged  $\alpha$ -diazocarbonyl compound 57, 25 mM of the azide and 10 mM sodium ascorbate in 100 mM MES buffer, pH 6.0 were kept at 20 °C for 24 h. Analysis of the reaction products was carried out by micropreparative HPLC-separation of 5- $\mu$ l aliquots of the reaction mixtures using method C. The collected peak fractions were further assessed by ESI-MS or MALDI TOF.



**Figure E29** | ESI-MS of +1<sub>Alk-Me</sub> (A), +1<sub>Alk/Click</sub> (B), +2<sub>Alk/Click</sub> (C), and +3<sub>Alk/Click</sub> (D) modified d(ATG) with diazo compound **57** and 3-azidoprop-1-ol.



**Figure E30** | MALDI TOF of the products from the tandem alkylation/'click' reaction of d(ATG C) with diazo compound **57** and 1-(2-azidoacetyl)-4-D-biotinylpiperazine (see Figure 24).



## 5.7. AlkB expression and purification

### 5.7.1. General

The restrictases *XhoI* and *NdeI* and T4 DNA ligase were from New England Biolabs, the restrictase *EcoRI* was from Roche. Kanamycin sulfate and IPTG were purchased from Fischer Biochemicals, Chloramphenicol was from Apollo. GenElute™ Plasmid Miniprep Kit was from Sigma, NucleoSpin® Gel and PCR Clean-up Kit was from Macherey-Nagel.

Protein SDS PAGE was carried out according to the method of Laemmli<sup>107</sup> in a BioRad MiniProtean Tetra system. Staining was done with Coomassie Brilliant Blue, the molecular weight of the proteins was referenced to Precision Plus Protein™ All Blue Standards (BioRad). DNA electrophoresis was done in a horizontal BioRad MiniSub Gel GT cell using 1 % (w/v) agarose (EEO-Mr < 0.15, Fischer Biochemicals) slabs and 1×TAE both for the gel and as electrode buffer. Bromphenol blue was used as a trace dye, visualization was done with ethidium bromide added to the gel slab. DNA size was referenced to 100 bp DNA ladder and 1 kb DNA ladder (New England Biolabs). Gel imaging was carried out on a BioRad Universal Hood III, equipped with Image Lab 5.0 software.

Plasmid sequencing was done by Microsynth AG.

### Culture media

LB: 10 g tryptone, 5 g yeast extract, 10 g NaCl, 1g agar (for solid media), water to 1000 ml.

Modified SOC: tryptone 8 g, yeast extract 2 g, NaCl 0.2 g, KCl 0.19 g, water to 400 ml; after autoclaving sterile-filtered 1M glucose (8 ml) was added.

Antibiotics were used in the following final concentrations: 50 µg/ml for kanamycin (Kan) from a sterile-filtered 1000× in water; 34 µg/ml for chloramphenicol (Cam) from a 1000× in ethanol.

### Buffers and solutions

50×TAE: 242 g Tris Base, 57.1 ml acetic acid, 100 ml 0.5M EDTA, pH 8.0, water to 1000 ml

IPTG – 200 mM in water, sterile-filtered

Lysis buffer: 500 mM NaCl, 50 mM sodium phosphate buffer, pH 8.0

Enzyme storage buffer: 300 mM NaCl, 2 mM DTT, 50 mM TrisHCl, pH 8.0

### Preparation of chemically competent *E.coli* cells

Competent *E.coli* NEB10beta and LB21 cells were prepared by treatment with CaCl<sub>2</sub> as follows. Single colonies were transferred in 3 ml LB and shaken overnight at 37°C. LB medium (100 ml) was inoculated with 1-ml of the overnight starter culture and grown at 37°C with shaking until OD<sub>600</sub> 0.40-0.45 was reached. The cultures were then cooled in ice, centrifuged (3000×g, 10 min, 4°C), and the cell pellets were drained over paper and resuspended in 20 ml ice-cold sterile 100 mM CaCl<sub>2</sub>. The

suspensions were left for 30 min in ice, then centrifuged, and resuspended in 6 ml ice-cold sterile 100 mM CaCl<sub>2</sub>, containing 15 % (v/v) glycerol. Fifty-microliter aliquots were then transferred to cold 1.5-ml tubes, flash-frozen in liquid nitrogen and stored at -80°C.

### **Heat shock transformation of chemically competent *E.coli* cells**

A fifty-microliter aliquot of the competent cells was mixed with plasmid DNA (10-100 ng) and left on ice for 30 min. The tubes were then incubated at 42°C for 1.5 min, and again on ice for 5 min. SOC medium (0.95 ml) was added to each tube, and the cultures were transferred into 15-ml sterile Falcon tubes and shaken at 37°C for one hour without addition of antibiotics. Aliquots (0.1 ml) were then plated on LB agar, containing the respective antibiotic for the selection, and the incubated overnight at 37°C. Single colonies of the transformants were then picked and grown in LB medium with the respective antibiotic.

### **5.7.2. Subcloning of the *alkB* gene from *E.coli* LB21**

The *alkB* gene was isolated from *E.coli* BL21 by colony PCR with Pfu+Taq polymerase mixture and the following thermocycler program min (°C): 5(95) ×1, 1(95), 1.5(54), 1(72) ×30, 10(72) ×1, 4(inf.).

The following primers were used:

- Forward primer: GGC AGC CAT ATG TTG GAT CTG TTT GCC GAT (T<sub>m</sub> 64°C (basic), GC 50%), contains restriction site for *NdeI*;
- Reverse primer: GTG GTG CTC GAG TTA TTC TTT TTT ACC TGC (T<sub>m</sub> 60°C (basic), GC 43%), contains restriction site for *XhoI*.

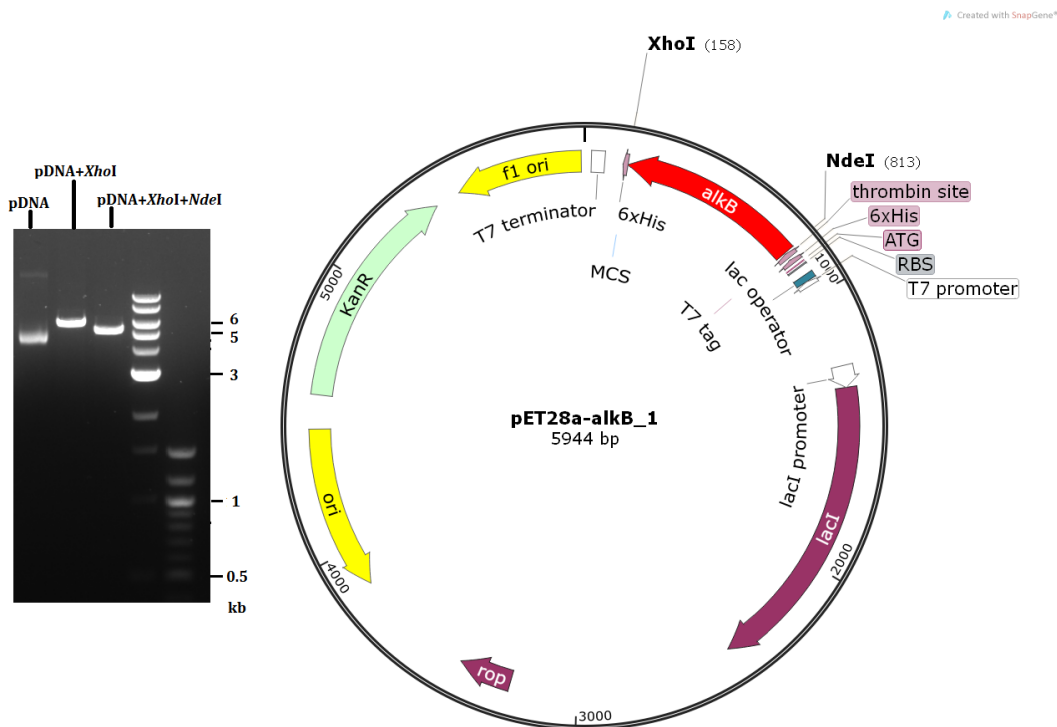
The PCR product was purified using the PCR purification kit (M.-N.), double digested with *NdeI* and *XhoI* according to the manufacturer's instructions. A larger amount of *NdeI* was used to insure full digestion. The obtained insert was ligated in a pET-28a vector using T4 DNA ligase and 100 ng vector in ratios vector to insert 1:3 and 1:6. Chemically competent *E.coli* NEB10beta cells (50 µl aliquots) were transformed with aliquots from the ligation mixtures (4 µl) and plated for selection on LB agar with Kan. Single colonies of the transformants were picked and grown in 10-ml LB medium with kanamycin for plasmid DNA isolation. The correct transformants were identified by restriction analysis and sequencing of the isolated plasmid DNA.

**Sequencing of the vector with the cloned insert (PCR-based sequencing with T7 promoter primer. The sequence for the N-terminal His<sub>6</sub>-tag is shown in grey)**

```

CTAGAAANATTTTGTTTACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATC
ATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGGATCTGTTTGCCGATG
GTGAACCGTGGCAAGAGCCACTGGCGGCTGGAGCGGTAATTTTACGGCGTTTTGCTTTTA
ACGCTGCGGAGCAACTAATCCGCGATATTAATGACGTTGCCAGCCAGTCGCCGTTTTCGCC
AGATGGTCACCCCCGGGGGATATACCATGTCTGGTGGCGATGACCAACTGTGGGCATCTGG
GCTGGACGAGCCATCGGCAAGGTTATCTCTATTCGCCATTGATCCGCAAACAAATAAAC
CGTGGCCCCGCCATGCCACAGAGTTTTTCATAATTTATGTCAACGTGCGGCTACGGCGGCGG
GCTATCCAGATTTCCAGCCAGATGCTTGTCTTATCAACCGCTACGCTCCTGGCGCGAAACT
GTCGCTGCATCAGGATAAAGACGAACCGGATCTGCGCGCGCCAATTGTTTCTGTTTCTCT
GGGCTTACCCGCGATTTTTCAATTTGGCGGCCTGAAACGTAATGATCCGCTCAAACGTTTG
TTGTTGGAACATGGCGATGTGGTGGTATGGGGCGGTGAATCGCGGCTGTTTTATCACGGT
ATTCAACCGTTGAAAGCGGGGTTTCATCCACTCACCACCGACTGCCGCTACAACCTGACA
TTCCGTCAGGCAGGTAAGAAAGAAATAACTCGAGCACCACCACCACCACCCTGAGATCC
GGCTGCTAACAAAGCCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAAC
TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAA
CTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGT
GGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGC
TTTCTTCCCTTCTTTCTCGCCACGTTCCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGC
TCCCTTAGGNTCCGATTNANNNNNTTACGGNNCCTCGACCCCAAAAACCTTGATTAGGN
NGA
  
```

**Theoretical size of the insert containing the *alkB* gene after double digestion: 655 bp**



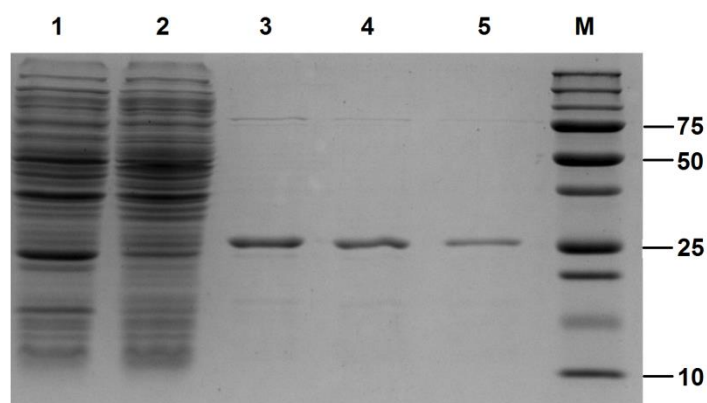
**Figure E31** | Agarose gel electrophoresis of the plasmid preparation from *E.coli* NEB10beta with restriction analysis (left), and map of the construct pET28a-*alkB* depicting the cloned insert (right). The construct map was prepared using SnapGene software.

**Expression product from the first reading frame (236 amino acids, theoretical MW 26198.74, in-frame with the His<sub>6</sub>-tag)**

10 20 30 40 50 60  
MGSSHHHHH SGLVPRGSH MLDLFDAGEP WQEPLAAGAV ILRRFAFNAA EQLIRDINDV  
70 80 90 100 110 120  
ASQSPFRQMV TPGGYTMSVA MTNCGHLGWT SHRQGYLYSP IDPQTNKPWP AMPQSFHNLK  
130 140 150 160 170 180  
QRAATAAGYP DFQPDACLIN RYAPGAKLSL HQDKDEPDLR APIVSVSLGL PAIFQFGGLK  
190 200 210 220 230  
RNDPLKRLLL EHGDIVVWGG ESRLFYHGIQ PLKAGFHPLT TDCRYNLTFR QAGKKE

**5.7.3. AlkB overexpression and purification**

Competent *E.coli* BL21 (50 µl aliquots) were transformed with the pET28a-*alkB* construct from 5.7.2 (50-100 ng per aliquot) by heat shock, then plated for selection on LB-agar+Kan+Cam. Single colonies were picked, transferred to 10 ml LB+Kan+Cam in 50 ml sterile Falcon tubes and grown overnight at room temperature. LB medium (1000 ml) was inoculated with 10 ml of the overnight starter culture and grown at 37°C until OD<sub>600</sub> of 0.7-0.8. The cultures were then briefly cooled in ice and induced with IPTG in a final concentration of 0.2 mM overnight at 18°C. The cells were harvested by centrifugation (3000×g, 20 min, 4°C), the cell pellets were resuspended in lysis buffer (1:5 v/v) and lysed by sonication. The cell lysates were clarified by centrifugation (4000×g, 20 min, 4°C), then mixed with Ni-NTA pre-washed with lysis buffer (ca. 0.1 ml of resin per liter of culture medium) and left for 20 min in ice with occasional swirling. The mixtures were then filtered on disposable plastic columns. The resin was washed with lysis buffer until A<sub>280</sub> of the effluent was smaller than 0.05, then with 10, 20, and 40 mM imidazole in lysis buffer (1 ml of each per 0.1 ml of resin). The target protein was eluted with 250 mM imidazole in lysis buffer (4 × 0.3 ml). The purity of the elution fractions was checked by SDS PAGE in 12.5 % resolving gel. After that the appropriate fractions were pooled, concentrated on a 3kD centrifuged filter (Amicon), and washed twice with ten volumes of enzyme storage buffer. Glycerol was then added to a final concentration of 50 % (v/v) from an 80 % (v/v) stock in water, and the preparations were stored at -20°C.



**Figure E32** | SDS PAGE analysis of the AlkB protein. Lane 1: clarified lysate, lane 2: flow through from loading the clarified lysate on the Ni-NTA resin, lanes 3 to 5: fractions from elution with 250 mM imidazole, M – molecular weight markers. The molecular weight of the standards are shown on the right.

#### 5.7.4. Determination of delacylation activity

Reaction mixtures (50  $\mu$ l) with the following content were prepared: 1 mM alkylated nucleoside monophosphate, 2 mM  $\alpha$ -ketoglutarate, 100  $\mu$ M FeSO<sub>4</sub>, 2 mM sodium ascorbate, 2 mM DTT, 50 mM KCl, 100  $\mu$ g.ml<sup>-1</sup> BSA, 50 mM Tris-HCl, pH 8.0, and 7.44  $\mu$ M AlkB. The mixtures were incubated at 37°C. Ten microliter aliquots were withdrawn at specific time intervals and mixed with 1  $\mu$ l 500 mM EDTA, pH 8.0, then analyzed by analytical RP-HPLC using method F. All reactions were done in duplicate together with control samples without enzyme. The reaction rates were determined from the disappearance of the nucleoside substrate or the accumulation of the repaired product, calculated from the peak areas in the HPLC-traces.

## References

---

- <sup>1</sup> A. Raj, P. van der Bogaard, S.A. Rifkin, A. van Oudenaarden, S. Tyagi, *Nat. Methods*, **2008**, 5, 877-879.
- <sup>2</sup> A. Tsourkas, M.A. Behlke, G. Bao, *Nucleic Acids Res.*, **2002**, 30, 4208-4215.
- <sup>3</sup> A. Tsourkas, M.A. Behlke, Y. Xu, G. Bao, *Anal. Chem.*, **2003**, 75, 3697-3703.
- <sup>4</sup> J. Ju, A.N. Glazer, R.A. Mathies, *Nucleic Acids Res.*, **1996**, 24, 1144-1148.
- <sup>5</sup> J. Gierlich, G.A. Burley, P.M.E. Gramlich, D.M. Hammond, T. Carrell, *Org. Lett.*, **2006**, 8, 3639-3642.
- <sup>6</sup> J. Gierlich, K. Gutsmedl, P.M.E. Gramlich, A. Schmidt, G.A. Burley, T. Carrell, *Chem. Eur. J.*, **2007**, 13, 9486-9494.
- <sup>7</sup> G.A. Burley, J. Gierlich, M.R. Mofid, H. Nir, S. Tal, Y. Eichen, T. Carrell, *J. Am. Chem. Soc.*, **2006**, 128, 1398-1399.
- <sup>8</sup> S.L. Beaucage, M.H. Caruthers, *Tetrahedron Lett.*, **1981**, 22, 1859-1862.
- <sup>9</sup> G.F. Kaufmann, M.M. Meijler, C.Sun, D.-W. Chen, D.P. Kujawa, J.M. Mee, T.Z. Hoffman, P. Wirsching, R.A. Lerner, K.D. Janda, *Angew. Chem. Int. Ed.*, **2005**, 44, 2144-2148.
- <sup>10</sup> Petr Capek, H. Cahov, R. Pohl, M. Hocek, C. Gloeckner, A. Marx, *Chem. Eur. J.*, **2007**, 13, 6196-6203.
- <sup>11</sup> S. Jaeger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.*, **127**, 15071-15082.
- <sup>12</sup> P.M.E. Gramlich, C.T. Wirges, J. Gierlich, T. Carrell, *Org. Lett.*, **2008**, 10, 249-251.
- <sup>13</sup> C.S. Show, S.K. Mahto, T.N. Lamichhane, *ACS Chem. Biol.*, **2008**, 3, 30-37.
- <sup>14</sup> N.E. Conway, L.W. McLaughlin, *Bioconjugate Chem.*, **1991**, 2, 452-457.
- <sup>15</sup> S.H. Weisbrod, A. Marx, *Chem. Commun.*, **2008**, 5675-5685.
- <sup>16</sup> S. Berndl, N. Herzig, P. Kele, D. Lachmann, X. Li, O.S. Wolfbeis, H.-A. Wagenknecht, *Bioconjugate Chem.*, **2009**, 20, 558-564.
- <sup>17</sup> K. Yamana, H. Zako, K. Asazuma, R. Iwase, H. Nakano, A. Murakami, *Angew. Chem. Int. Ed.*, **2001**, 40, 1104-1106.
- <sup>18</sup> F. Seela, M. Zulauf, *Chem. Eur. J.*, **1998**, 4, 1781-1790.
- <sup>19</sup> R.T. Ranasinghe, T. Brown, *Chem. Commun.*, **2005**, 5487-5502.
- <sup>20</sup> A.H. El-Sagheer, A.P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *PNAS*, **2011**, 108, 11338-11343.
- <sup>21</sup> R. Roychoudhury, E. Jay, R. Wu, *Nucleic Acids Res.*, **1976**, 3, 863-877.
- <sup>22</sup> C.E. Guerra, *BioTechniques*, **2006**, 41, 53-56.
- <sup>23</sup> G.L. Igloi, *BioTechniques*, **1996**, 21, 1084-1092.
- <sup>24</sup> G. Martin, W. Keller, *RNA*, **1998**, 4, 226-230.
- <sup>25</sup> M.J. Moore, P.A. Sharp, *Science*, **1992**, 256, 992-997.

- 
- <sup>26</sup> M.-L. Winz, A. Samanta, D. Benzinger, A. Jaeschke, *Nucleic Acids Res.*, **2012**, 40, e78.
- <sup>27</sup> Y. Tor, P.B. Dervan, *J. Am. Chem. Soc.*, **1993**, 115, 4461-4467.
- <sup>28</sup> R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc.*, **2005**, 127, 17286-17295.
- <sup>29</sup> K. Moriyama, M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.*, **2005**, 33, e129.
- <sup>30</sup> G. Pljevaljcic, M. Pignot, E. Weinhold, *J. Am. Chem. Soc.*, **2003**, 125, 3486-3492.
- <sup>31</sup> G. Pljevaljc, F. Schmidt, E. Weinhold, *ChemBioChem*, **2004**, 5, 265-269.
- <sup>32</sup> G. Pljevaljc, F. Schmidt, A.J. Scheidig, R. Lurz, E. Weinhold, *ChemBioChem*, **2007**, 8, 1516-1519.
- <sup>33</sup> F.H.-G. Schmidt, M. Hüben, B. Gider, F. Renault, M.-P. Teulade-Fichou, E. Weinhold, *Bioorg. Med. Chem.*, **2008**, 16, 40-48.
- <sup>34</sup> G. Lukinavicius, V. Lapiene, Z. Stasevskij, C. Dalhoff, E. Weinhold, S. Klimasauskas, *J. Am. Chem. Soc.*, **2007**, 129, 2758-2759.
- <sup>35</sup> Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold, M. Helm, *Nucleic Acids Res.*, **2011**, 39, 1943-1952.
- <sup>36</sup> D.E. Hatchway, G.F. Kolar, *Chem. Soc. Rev.*, **1980**, 9, 241-264.
- <sup>37</sup> Y. Mishina, E.M. Duguid, C. He, *Chem. Rev.*, **2006**, 106, 215-232.
- <sup>38</sup> V. Monnot, C. Tora, S. Lopez, L. Menou, A. Laayoun, *Nucleosides Nucleotides Nucleic Acids*, **2001**, 20, 1177-1179.
- <sup>39</sup> A. Laayoun, M. Kotera, I. Sothier, E. Trévisiol, E. Bernal-Méndez, C. Bourget, L. Menou, J. Lhomme, A. Troesch, *Bioconjugate Chem.*, **2003**, 14, 1298-1306.
- <sup>40</sup> K. Stevens, A. Madder, *Nucleic Acids Research*, **2009**, 37, 1555-1565.
- <sup>41</sup> M. Op de Beeck, A. Madder, *J. Am. Chem. Soc.*, **2011**, 133, 769-807.
- <sup>42</sup> K. Stevens, D.D. Claeys, S. Catak, S. Figaroli, M. Hocek, J.M. Tromp, S. Schürch, V. Van Speybroeck, A. Madder, *Chem. Eur. J.*, **2011**, 17, 6940-6953.
- <sup>43</sup> M. Op de Beeck, A. Madder, *J. Am. Chem. Soc.*, **2012**, 134, 10737-10740.
- <sup>44</sup> P. Pande, J. Shearer, J. Yang, W.A. Greenberg, S.E. Rokita, *J. Am. Chem. Soc.*, **1999**, 121, 6773-6779.
- <sup>45</sup> Q. Zhou, S. Rokita, *PNAS*, **2003**, 100, 15452-15457.
- <sup>46</sup> K. Onizuka, Y. Taniguchi, S. Sasaki, *J. Am. Chem. Soc.*, **2009**, 131, 799-803.
- <sup>47</sup> G. Lukinavičius, A. Lapinaitė, G. Urbanavičiūtė, R. Gerasimaitė, S. Klimašauskas, *Nucleic Acids Res.*, **2012**, 40, 11594-11602.
- <sup>48</sup> T. Ye, M.A. McKervey, *Chem. Rev.*, **1994**, 94, 1091-1160.
- <sup>49</sup> E. Wenkert, L.L. Davis, B.L. Mylari, M.F. Solomon, R.R. Da Silva, S. Shulman, R.J. Warnet, P. Ceccherelli, M. Curini, R. Pellicciari, *J. Org. Chem.*, **1982**, 47, 3242-3247.
- <sup>50</sup> V. Bagheri, M.P. Doyle, J. Taunton, E.E. Claxton, *J. Org. Chem.*, **1988**, 53, 6158.

- 
- <sup>51</sup> A.F. Noels, A. Demonceau, N. Petiniot, A.J. Hubert, Ph. Teyssié, *Tetrahedron*, **1982**, 38, 2733-2739.
- <sup>52</sup> T.N. Salzmann, R.W. Ratcliffe, B.G. Christensen, F.A. Bouffard, *J. Am. Chem. Soc.*, **1980**, 102, 6161-6163.
- <sup>53</sup> M.A. McKerverey, P. Ratananukul, *Tetrahedron Lett.*, **1982**, 23, 2509-2512.
- <sup>54</sup> J.M. Antos, M.B. Francis, *J. Am. Chem. Soc.*, **2004**, 126, 10256-10257.
- <sup>55</sup> J.M. Antos, J.M. McFarland, A.T. Iavarone, M.B. Francis, *J. Am. Chem. Soc.*, **2009**, 131, 6301-6308.
- <sup>56</sup> B.V. Popp, Z.T. Ball, *J. Am. Chem. Soc.*, **2010**, 132, 6660-6662.
- <sup>57</sup> B.V. Popp, Z.T. Ball, *Chem. Sci.*, **2011**, 2, 690-695.
- <sup>58</sup> Z. Chen, B.V. Popp, C.L. Bovet, Z.T. Ball, *ACS Chem. Biol.*, **2011**, 6, 920-925.
- <sup>59</sup> P. Yates, *J. Am. Chem. Soc.*, **1952**, 74, 5376-5281.
- <sup>60</sup> A.J. Anciaux, A.J. Hubert, A.F. Noels, N. Petiniot, P. Teyssie, *J. Org. Chem.*, **1980**, 45, 695-702.
- <sup>61</sup> M.C. Purring, A.T. Morehead, *J. Am. Chem. Soc.*, **1996**, 118, 8162-81-63.
- <sup>62</sup> M.C. Purring, H. Liu, A.T. Morehead, *J. Am. Chem. Soc.*, **2002**, 124, 1014-1023.
- <sup>63</sup> D.T. Nowlan, T.M. Gregg, H.M.L. Davies, D.A. Singleton, *J. Am. Chem. Soc.*, **2003**, 125, 15902-15911.
- <sup>64</sup> E. Nakamura, N. Yoshikai, M. Yamanaka, *J. Am. Chem. Soc.*, **2002**, 124, 7181-7192.
- <sup>65</sup> M.P. Doyle, L.J. Westrum, W.N. E. Wolthuis, M.M. See, W.P. Boone, V. Bagheri, M.M. Pearson, *J. Am. Chem. Soc.*, **1993**, 115, 958-964.
- <sup>66</sup> D.F. Taber, K.K. You, A.L. Rheingold, *J. Am. Chem. Soc.*, **1996**, 118, 547-556.
- <sup>67</sup> D.F. Taber, E.H. Petty, K. Raman, *J. Am. Chem. Soc.*, **1985**, 107, 196-199.
- <sup>68</sup> P. Bulugahapitiya, Y. Landais, L. Parra-Rapado, D. Planchenault, V. Weber, *J. Org. Chem.*, **1997**, 62, 1630-1641.
- <sup>69</sup> M.P. Doyle, M. Yan, *Tetrahedron Lett.*, **2002**, 43, 5929-5931.
- <sup>70</sup> Z. Qu, W. Shi, J. Wang, *J. Org. Chem.*, **2004**, 69, 217-219.
- <sup>71</sup> F.A. Davis, B. Yang, J. Deng, *J. Org. Chem.*, **2003**, 68, 5147-5152.
- <sup>72</sup> H. Huang, X. Guo, W. Hu, *Angew. Chem. Int. Ed.*, **2007**, 46, 1337-1339.
- <sup>73</sup> Z. Li, B.T. Parr, H.M.L. Davies, *J. Am. Chem. Soc.*, **2012**, 134, 10942-10946.
- <sup>74</sup> S. Bachmann, D. Fielenbach, K.A. Jorgensen, *Org. Biomol. Chem.*, **2004**, 2, 3044-3049.
- <sup>75</sup> E.C. Lee, G.C. Fu, *J. Am. Chem. Soc.*, **2007**, 129, 12066-12067.
- <sup>76</sup> T.C. Maier, G.C. Fu, *J. Am. Chem. Soc.*, **2006**, 128, 4594-4595.
- <sup>77</sup> S.-F. Zhou, B. Xu, G.-P. Wang, Q.-L. Zhou, *J. Am. Chem. Soc.*, **2012**, 134, 436-442.
- <sup>78</sup> R.G. Salomon, J.K. Kochi, *J. Am. Chem. Soc.*, **1973**, 95, 3300-3310.
- <sup>79</sup> Y. Liang, H. Zhou, Z.-X. Yu, *J. Am. Chem. Soc.*, **2009**, 131, 17783-17785.



- 
- <sup>80</sup> V.V. Rostovtsev, L.G. Green, V.V. Fokin, K.B. Sharpless, *Angew. Chem. Int. Ed.*, **2002**, 41, 2596-2599.
- <sup>81</sup> C.W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.*, **2002**, 67, 3057-3062.
- <sup>82</sup> H.C. Kolb, M.G. Finn, K.B. Sharpless, *Angew. Chem. Int. Ed.*, **2001**, 40, 2004-2021.
- <sup>83</sup> A.H. El-Sagheer, T. Brown, *Chem. Soc. Rev.*, **2010**, 39, 1388-1405.
- <sup>84</sup> S.K. Mamidyala, M.G. Finn, *Chem. Soc. Rev.*, **2010**, 39, 1252-1261.
- <sup>85</sup> J.E. Hein, V.V. Fokin, *Chem. Soc. Rev.*, **2010**, 39, 1302-1315.
- <sup>86</sup> V. Hong, S.I. Presolski, C. Ma, M.G. Finn, *Angew. Chem. Int. Ed.*, **2009**, 48, 9879-9883.
- <sup>87</sup> T.R. Chan, R. Hilgraf, K.B. Sharpless, V.V. Fokin, *Org. Lett.*, **2004**, 6, 2853-2855.
- <sup>88</sup> B.T. Worrell, J.A. Malik, V.V. Fokin, *Science*, **2013**, 340, 457-460.
- <sup>89</sup> H.T. Chifotides, K.R. Dunbar, *Acc. Chem. Res.*, **2005**, 38, 146-156.
- <sup>90</sup> H.T. Chifotides, K.M. Koshlap, L.M. Perez, K.R. Dunbar, *J. Am. Chem. Soc.*, **2003**, 125, 10703-10713.
- <sup>91</sup> A.M. Angeles-Boza, H.T. Chifotides, J.D. Aguirre, A. Chouai, P.K.-L. Fu, K.R. Dunbar, C. Turro, *C. J. Med. Chem.* **2006**, 49, 6841-6847.
- <sup>92</sup> J.D. Aguirre, A.M. Angeles-Boza, A. Chouai, J.-P. Pellois, C. Turro, K.R. Dunbar, *J. Am. Chem. Soc.* **2009**, 131, 11353-11360.
- <sup>93</sup> N. Farrell, *J. Inorg. Biochem.*, **1981**, 14, 261-265.
- <sup>94</sup> B. Zümreoglu-Karan, *Coord. Chem. Rev.*, **2006**, 250, 2295-2307.
- <sup>95</sup> R.O. Clinton, S.C. Laskowski, *J. Am. Chem. Soc.*, **1948**, 70, 3135-3136.
- <sup>96</sup> P. Koivisto, T. Duncan, T. Lindahl, B. Sedgwick, *J. Biol. Chem.*, **2003**, 278, 44348-44354.
- <sup>97</sup> J.C. Delaney, J.M. Essigmann, *PNAS*, **2004**, 101, 14051-14056.
- <sup>98</sup> A.M. Maciejewska, K.P. Ruszel, J. Nieminuszczy, J. Lewicka, B. Sokołowska, E. Grzesiuk, J.T. Kusmierek, *Mutat. Res.*, **2010**, 684, 24-34.
- <sup>99</sup> R. Ottria, S. Casati, A. Manzocchi, E. Baldoli, M. Mariotti, J.A.M. Maier, P. Ciuffreda, *Bioorg. Med. Chem.*, **2010**, 18, 4249-4254.
- <sup>100</sup> S.C. Tregwick, T.F. Henshaw, R.P. Hausinger, T. Lindahl, B. Sedgwick, *Nature*, **2002**, 419, 174-178.
- <sup>101</sup> D. Liu, L. A. Gugliotti, T. Wu, M. Dolska, A. Tkachenko, M. K. Shipton, B. E. Eaton, and D. L. Feldheim, *Langmuir*, **2006**, 22, 5862-5866.
- <sup>102</sup> L. E. Smeenk, N. Dailly, H. Hiemstra, J. van Maarseveen, and P. Timmerman, *Org. Lett.*, **2012**, 14, 1194-1197.
- <sup>103</sup> T. L. Hwang, A. J. Shaka, *J. Magn. Reson.*, **1995**, A115, 275-279.
- <sup>104</sup> T. Peters Jr., *The Plasma Proteins*. F.W. Putnam, ed. Academic Press., **1975**, pp. 133-181.

- 
- <sup>105</sup> Worthington Enzyme Manual. K. Worthington, V. Worthington, 2011, Worthington Biochemical Corporation. June, 6<sup>th</sup> **2013** (<http://www.worthington-biochem.com/LY/default.html>).
- <sup>106</sup> S.C. Gill, P.H. von Hippel, *Anal. Biochem.*, **1989**, 182, 319-326.
- <sup>107</sup> U.K. Laemmli, *Nature*, **1970**, 227, 680-685.

## Kiril Tishinov

---

[ktishinov@gmail.com](mailto:ktishinov@gmail.com)

11.08.1981  
Bulgarian

---

### Education

---

**02/2011-12/2014**

**PhD Degree (*summa cum laude*) in Chemistry** in the group of Prof. Dennis Gillingham, Department of Chemistry, **University of Basel**

Topic: *'New catalytic strategies for DNA and RNA alkylation using Rh(II) and Cu(I) carbenes – a versatile tool for applications in chemical biology'*

**11/2006-03/2008**

**Master's Degree in Biophysical Chemistry in Medicine and Pharmacology**, Faculty of Chemistry, **Sofia University**

**11/2004-03/2008**

**Master's Degree in Biotechnologies**, Faculty of Chemical and System Engineering, **University of Chemical Technology and Metallurgy, Sofia**

**10/2000-07/2004**

**Bachelor's Degree in Molecular Biology**, Faculty of Biology, **Sofia University**

---

### Internships

---

**06-08/2008 and 10-12/2007**

**DFG-Supported internships** in the group of Prof. Andreas Liese, Institute of Technical Biocatalysis, **Hamburg University of Technology**

---

### Working experience

---

**02/2011-present**

**Teaching Assistant**, Bachelor practical courses in Organic Chemistry, University of Basel

**06/2008-10/2010**

**Junior scientist**, Institute of Organic Chemistry, Sofia

**10/2004-06/2008**

**Research assistant**, Institute of Organic Chemistry, Sofia

---

### Professional expertise

---

- Synthetic organic chemistry
  - Solid-phase oligonucleotide synthesis
  - Analytical and preparative HPLC-techniques, including LC-MS
  - Mass spectrometry (ESI-MS and MS/MS, MALDI-TOF)
  - 1D- and 2D-NMR in solution
  - Protein expression and purification
  - Enzyme kinetics
-

## Selected publications

---

- (1) P. Schmidt, L. Zhou, **K. Tishinov**, K. Zimmermann, D. Gillingham. Dialdehydes lead to exceptionally fast bioconjugations at neutral pH by virtue of a cyclic intermediate. *Angew. Chem. Int. Ed.*, **2014**, 53, 10928-10931.
  - (2) **K. Tishinov**, N. Fei, D. Gillingham. Cu(I)-Catalyzed NH-insertion in water: a new tool for chemical biology, *Chem. Sci.*, **2013**, 4, 4401.
  - (3) D. Gillingham, **K. Tishinov**. Synthesis of nucleic acid polymers with non-canonical bases, *Synlett*, **2013**, 24, 893.
  - (4) **K. Tishinov**, K. Schmidt, D. Haeussinger, D. Gillingham. Structure-selective catalytic alkylation of DNA and RNA, *Angew. Chem. Int. Ed.*, **2012**, 51, 12000.
  - (5) **K. Tishinov**, S. Bayryamov, P. Nedkov, N. Stambolieva, B. Galunsky. A highly enantioselective aminopeptidase from sunflower seed—kinetic studies, substrate mapping and application to biocatalytic transformations, *J. Mol. Catal. B: Enzymatic*, **2009**, 59, 106.
  - (6) **K. Tishinov**, N. Stambolieva, S. Petrova, B. Galunsky, P. Nedkov. Purification and characterization of the sunflower seed (*Helianthus annuus* L.) major aminopeptidase, *Acta Physiol. Plant.*, **2009**, 31, 199.
- 

## Participation in scientific forums

---

- (1) **K. Tishinov**, D. Gillingham. Catalytic Alkylation of Nucleic Acids (*oral presentation*). 4th International Symposium on DNA-Encoded Chemical Libraries, Zurich, Switzerland, 2014.
  - (2) **K. Tishinov**, D. Gillingham. Catalytic Strategies for Nucleic Acid Alkylation (*poster*). Swiss Chemical Society Fall Meeting, Lausanne, Switzerland, 2013.
- 

## Professional associations

---

Member of the Swiss Chemical Society  
Member of the American Chemical Society

---

## Language proficiency

---

**English** (fluent)  
**German** (full professional proficiency)  
**Bulgarian** (fluent, mother tongue)

---

## Interests and hobbies

---

Skiing, climbing, bouldering,  
trekking, cooking

---